

Robust adaptive efficient estimation for semi-Markov nonparametric regression models *

Vlad Stefan Barbu[†], Slim Beltaief[‡] and Serguei Pergamenshchikov[§]

Abstract

We consider the nonparametric robust estimation problem for regression models in continuous time with semi-Markov noises. An adaptive model selection procedure is proposed. Under general moment conditions on the noise distribution a sharp non-asymptotic oracle inequality for the robust risks is obtained and the robust efficiency is shown. It turns out that for semi-Markov models the robust minimax convergence rate may be faster or slower than the classical one.

MSC: primary 62G08, secondary 62G05

Keywords: Non-asymptotic estimation; Robust risk; Model selection; Sharp oracle inequality; Asymptotic efficiency.

*This work was done under partial financial support of the grant of RSF number 14-49-00079 (National Research University “MPEI” 14 Krasnokazarmennaya, 111250 Moscow, Russia) and of the RFBR Grant 16-01-00121.

[†]Laboratoire de Mathématiques Raphaël Salem, UMR 6085 CNRS-Université de Rouen, France, e-mail: barbu@univ-rouen.fr

[‡]Laboratoire de Mathématiques Raphaël Salem, UMR 6085 CNRS-Université de Rouen, France, e-mail: slim.beltaief1@univ-rouen.fr

[§]Laboratoire de Mathématiques Raphaël Salem, UMR 6085 CNRS-Université de Rouen, France and International Laboratory of Statistics of Stochastic Processes and Quantitative Finance of National Research Tomsk State University, e-mail: Serge.Pergamenshchikov@univ-rouen.fr

1 Introduction

Let us consider a regression model in continuous time

$$dy_t = S(t)dt + d\xi_t, \quad 0 \leq t \leq n, \quad (1.1)$$

where $S(\cdot)$ is an unknown 1-periodic function from $\mathbf{L}_2[0, 1]$ defined on \mathbb{R} with values in \mathbb{R} , the noise process $(\xi_t)_{t \geq 0}$ is defined as

$$\xi_t = \varrho_1 L_t + \varrho_2 z_t, \quad (1.2)$$

where ϱ_1 and ϱ_2 are unknown coefficients, $(L_t)_{t \geq 0}$ is a Levy process and the pure jump process $(z_t)_{t \geq 1}$, defined in (2.3), is assumed to be a semi-Markov process (see, for example, [2]).

The problem is to estimate the unknown function S in the model (1.1) on the basis of observations $(y_t)_{0 \leq t \leq n}$. Firstly, this problem was considered in the framework of the “signal+white noise” models (see, for example, [6] or [24]). Later, in order to study dependent observations in continuous time, were introduced “signal+color noise” regressions based on Ornstein-Uhlenbeck processes (cf. [8], [9], [10], [13]).

Moreover, to include jumps in such models, the papers [14] and [15] used non Gaussian Ornstein-Uhlenbeck processes introduced in [1] for modeling of the risky assets in the stochastic volatility financial markets. Unfortunately, the dependence of the stable Ornstein-Uhlenbeck type decreases with a geometric rate. So, asymptotically when the duration of observations goes to infinity, we obtain very quickly the same “signal+white noise” model.

The main goal of this paper is to consider continuous time regression models with dependent observations for which the dependence does not disappear for a sufficient large duration of observations. To this end we define the noise in the model (1.1) through a semi-Markov process which keeps the dependence for any duration n . This type of models allows, for example, to estimate the signals observed under long impulse noise impact with a memory or “against signals”.

In this paper we use the robust estimation approach introduced in [14] for such problems. To this end, we denote by Q the distribution of $(\xi_t)_{0 \leq t \leq n}$ in the Skorokhod space $\mathcal{D}[0, n]$. We assume that Q is unknown and belongs to some distribution family \mathcal{Q}_n specified in Section 4. In this paper we use the quadratic risk

$$\mathcal{R}_Q(\tilde{S}_n, S) = \mathbf{E}_{Q,S} \|\tilde{S}_n - S\|^2, \quad (1.3)$$

where $\|f\|^2 = \int_0^1 f^2(s)ds$ and $\mathbf{E}_{Q,S}$ is the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process (1.1) corresponding to the noise distribution Q . Since

the noise distribution Q is unknown, it seems reasonable to introduce the robust risk of the form

$$\mathcal{R}_n^*(\tilde{S}_n, S) = \sup_{Q \in \mathcal{Q}_n} \mathcal{R}_Q(\tilde{S}_n, S), \quad (1.4)$$

which enables us to take into account the information that $Q \in \mathcal{Q}_n$ and ensures the quality of an estimate \tilde{S}_n for all distributions in the family \mathcal{Q}_n .

To summarize, the goal of this paper is to develop robust efficient model selection methods for the model (1.1) with the semi-Markov noise having unknown distribution, based on the approach proposed by Konev and Pergamenshchikov in [14] and [15] for continuous time regression models with semimartingale noises. Unfortunately, we cannot use directly this method for semi-Markov regression models, since their tool essentially uses the fact that the Ornstein-Uhlenbeck dependence decreases with geometrical rate and the “white noise” case is obtained sufficiently quickly.

Thus in the present paper we propose new analytical tools based on renewal methods to obtain the sharp non-asymptotic oracle inequalities. As a consequence, we obtain the robust efficiency for the proposed model selection procedures in the adaptive setting.

The rest of the paper is organized as follows. We start by introducing the main conditions in the next section. Then, in Section 3 we construct the model selection procedure on the basis of the weighted least squares estimates. The main results are stated in Section 4; here we also specify the set of admissible weight sequences in the model selection procedure. In Section 5 we derive some renewal results useful for obtaining other results of the paper. In Section 6 we develop stochastic calculus for semi Markov processes. In Section 7 we study some properties of the model (1.1). A numerical example is presented in Section 8. Most of the results of the paper are proved in Section 9. In Appendix some auxiliary propositions are given.

2 Main conditions

In the model (1.2) we assume that the Levy process L_t is defined as

$$L_t = \check{\varrho} w_t + \sqrt{1 - \check{\varrho}^2} \check{L}_t, \quad \check{L}_t = x * (\mu - \tilde{\mu})_t, \quad (2.1)$$

where, $0 \leq \check{\varrho} \leq 1$ is an unknown constant, $(w_t)_{t \geq 0}$ is a standard Brownian motion, $\mu(ds, dx)$ is the jump measure with the deterministic compensator $\tilde{\mu}(ds dx) = ds \Pi(dx)$, where $\Pi(\cdot)$ is some positive measure on \mathbb{R} (see, for example [7, 3] for details) for which we assume that

$$\Pi(x^2) = 1 \quad \text{and} \quad \Pi(x^8) < \infty, \quad (2.2)$$

where we use the usual notation $\Pi(|x|^m) = \int_{\mathbb{R}} |z|^m \Pi(dz)$ for any $m > 0$. Note that $\Pi(\mathbb{R})$ may be equal to $+\infty$. Moreover, we assume that the pure jump process $(z_t)_{t \geq 0}$ in (1.2) is a semi-Markov process with the following form

$$z_t = \sum_{i=1}^{N_t} Y_i, \quad (2.3)$$

where $(Y_i)_{i \geq 1}$ is an i.i.d. sequence of random variables with

$$\mathbf{E} Y_i = 0, \quad \mathbf{E} Y_i^2 = 1 \quad \text{and} \quad \mathbf{E} Y_i^4 < \infty.$$

Here N_t is a general counting process (see, for example, [19]) defined as

$$N_t = \sum_{k=1}^{\infty} \mathbb{1}_{\{T_k \leq t\}} \quad \text{and} \quad T_k = \sum_{l=1}^k \tau_l, \quad (2.4)$$

where $(\tau_l)_{l \geq 1}$ is an i.i.d. sequence of positive integrated random variables with distribution η and mean $\tilde{\tau} = \mathbf{E} \tau_1 > 0$. We assume that the processes $(N_t)_{t \geq 0}$ and $(Y_i)_{i \geq 1}$ are independent between them and are also independent of $(L_t)_{t \geq 0}$.

Note that the process $(z_t)_{t \geq 0}$ is a special case of a semi-Markov process (see, e.g., [2] and [17]).

Remark 2.1. *It should be noted that if τ_j are exponential random variables, then $(N_t)_{t \geq 0}$ is a Poisson process and, in this case, $(\xi_t)_{t \geq 0}$ is a Levy process for which this model has been studied in [11], [12] and [14]. But, in the general case when the process (2.3) is not a Levy process, this process has a memory and cannot be treated in the framework of semi-martingales with independent increments. In this case, we need to develop new tools based on renewal theory arguments, what we do in Section 5. This tools will be intensively used in the proofs of the main results of this paper.*

Note that for any function f from $\mathbf{L}_2[0, n]$, $f : [0, n] \rightarrow \mathbb{R}$, for the noise process $(\xi_t)_{t \geq 0}$ defined in (1.2), with $(z_t)_{t \geq 0}$ given in (2.3), the integral

$$I_n(f) = \int_0^n f(s) d\xi_s \quad (2.5)$$

is well defined with $\mathbf{E}_Q I_n(f) = 0$. Moreover, as it is shown in Lemma 6.2,

$$\mathbf{E}_Q I_n^2(f) \leq \varkappa_Q \int_0^n f^2(s) ds, \quad (2.6)$$

where $\varkappa_Q = \varrho_1^2 + \varrho_2^2 |\rho|_*$ and $|\rho|_* = \sup_{t \geq 0} |\rho(t)| < \infty$. Here ρ is the density of the renewal measure $\check{\eta}$ defined as

$$\check{\eta} = \sum_{l=1}^{\infty} \eta^{(l)}, \quad (2.7)$$

where $\eta^{(l)}$ is the l th convolution power for η . To study the series (2.7) we assume that the measure η has a density g which satisfies the following conditions.

H₁) Assume that, for any $x \in \mathbb{R}$, there exist the finite limits

$$g(x-) = \lim_{z \rightarrow x-} g(z) \quad \text{and} \quad g(x+) = \lim_{z \rightarrow x+} g(z)$$

and, for any $K > 0$, there exists $\delta = \delta(K) > 0$ for which

$$\sup_{|x| \leq K} \int_0^\delta \frac{|g(x+t) + g(x-t) - g(x+) - g(x-)|}{t} dt < \infty.$$

H₂) For any $\gamma > 0$,

$$\sup_{z \geq 0} z^\gamma |2g(z) - g(z-) - g(z+)| < \infty.$$

H₃) There exists $\beta > 0$ such that $\int_{\mathbb{R}} e^{\beta x} g(x) dx < \infty$.

Remark 2.2. It should be noted that the condition **H₃**) means that there exists an exponential moment for the random variable $(\tau_j)_{j \geq 1}$, i.e. these random variables are not too large. This is a natural constraint since these random variables define the intervals between jumps, i.e., the frequency of the jumps. So, to study the influence of the jumps in the model (1.1) one needs to consider the noise process (1.2) with “small” interval between jumps or large jump frequency.

For the next condition we need to introduce the Fourier transform of any function f from $\mathbf{L}_1(\mathbb{R})$, $f : \mathbb{R} \rightarrow \mathbb{R}$, defined as

$$\widehat{f}(\theta) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\theta x} f(x) dx. \quad (2.8)$$

H₄) There exists $t^* > 0$ such that the function $\widehat{g}(\theta - it)$ belongs to $\mathbf{L}_1(\mathbb{R})$ for any $0 \leq t \leq t^*$.

It is clear that Conditions $\mathbf{H}_1)$ – $\mathbf{H}_4)$ hold true for any continuously differentiable function g , for example for the exponential density.

Now we define the family of the noise distributions for the model (1.1) which is used in the robust risk (1.4). Note that any distribution Q from \mathcal{Q}_n is defined by the unknown parameters in (1.2) and (2.1). We assume that

$$\varsigma_* \leq \varrho_1^2 \leq \varsigma^*, \quad 0 \leq \varrho \leq 1 \quad \text{and} \quad \varsigma_* \leq \sigma_Q \leq \varsigma^*, \quad (2.9)$$

where $\sigma_Q = \varrho_1^2 + \varrho_2^2/\tilde{\tau}$, the unknown bounds $0 < \varsigma_* \leq \varsigma^*$ are functions of n , i.e. $\varsigma_* = \varsigma_*(n)$ and $\varsigma^* = \varsigma^*(n)$, such that for any $\tilde{\epsilon} > 0$,

$$\lim_{n \rightarrow \infty} n^{\tilde{\epsilon}} \varsigma_*(n) = +\infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{\varsigma^*(n)}{n^{\tilde{\epsilon}}} = 0. \quad (2.10)$$

Remark 2.3. As we will see later, the parameter σ_Q is the limit for the Fourier transform of the noise process (1.2). Such limit is called variance proxy (see [14]).

Remark 2.4. Note that, generally (but it is not necessary) the parameters ϱ_1 and ϱ_2 can be dependent on n . The conditions (2.10) means that we consider all possible cases, i.e. these parameters may go to the infinity or be constant or to the zero as well. See, for example, the conditions (3.32) in [15].

3 Model selection

Let $(\phi_j)_{j \geq 1}$ be an orthonormal uniformly bounded basis in $\mathbf{L}_2[0, 1]$, i.e., for some constant $\phi_* \geq 1$, which may be depend on n ,

$$\sup_{0 \leq j \leq n} \sup_{0 \leq t \leq 1} |\phi_j(t)| \leq \phi_* < \infty. \quad (3.1)$$

We extend the functions $\phi_j(t)$ by periodicity, i.e., we set $\phi_j(t) := \phi_j(\{t\})$, where $\{t\}$ is the fractional part of $t \geq 0$. For example, we can take the trigonometric basis defined as $\text{Tr}_1 \equiv 1$ and, for $j \geq 2$,

$$\text{Tr}_j(x) = \sqrt{2} \begin{cases} \cos(2\pi[j/2]x) & \text{for even } j; \\ \sin(2\pi[j/2]x) & \text{for odd } j, \end{cases} \quad (3.2)$$

where $[x]$ denotes the integer part of x .

To estimate the function S we use here the model selection procedure for continuous time regression models from [14] based on the Fourier expansion. We recall that for any function S from $\mathbf{L}_2[0, 1]$ we can write

$$S(t) = \sum_{j=1}^{\infty} \theta_j \phi_j(t) \quad \text{and} \quad \theta_j = (S, \phi_j) = \int_0^1 S(t) \phi_j(t) dt. \quad (3.3)$$

So, to estimate the function S it suffices to estimate the coefficients θ_j and to replace them in this representation by their estimators. Using the fact that the function S and ϕ_j are 1 - periodic we can write that

$$\theta_j = \frac{1}{n} \int_0^n \phi_j(t) S(t) dt .$$

If we replace here the differential $S(t)dt$ by the stochastic observed differential dy_t we obtain the natural estimate for θ_j on the time interval $[0, n]$

$$\hat{\theta}_{j,n} = \frac{1}{n} \int_0^n \phi_j(t) dy_t , \quad (3.4)$$

which can be represented, in view of the model (1.1), as

$$\hat{\theta}_{j,n} = \theta_j + \frac{1}{\sqrt{n}} \xi_{j,n} , \quad \xi_{j,n} = \frac{1}{\sqrt{n}} I_n(\phi_j) . \quad (3.5)$$

Now (see, for example, [6]) we can estimate the function S by the projection estimators, i.e.

$$\hat{S}_m(t) = \sum_{j=1}^m \hat{\theta}_{j,n} \phi_j(t) , \quad 0 \leq t \leq 1 , \quad (3.6)$$

for some number $m \rightarrow \infty$ as $n \rightarrow \infty$. It should be noted that Pinsker in [24] shows that the projection estimators of the form (3.6) are not efficient. For obtaining efficient estimation one needs to use weighted least square estimators defined as

$$\hat{S}_\lambda(t) = \sum_{j=1}^n \lambda(j) \hat{\theta}_{j,n} \phi_j(t) , \quad (3.7)$$

where the coefficients $\lambda = (\lambda(j))_{1 \leq j \leq n}$ belong to some finite set Λ from $[0, 1]^n$. As it is shown in [24], in order to obtain efficient estimators, the coefficients $\lambda(j)$ in (3.7) need to be chosen depending on the regularity of the unknown function S . In this paper we consider the adaptive case, i.e. we assume that the regularity of the function S is unknown. In this case we chose the weight coefficients on the basis of the model selection procedure proposed in [14] for the general semi-martingale regression model in continuous time. These coefficients will be obtained later in (3.20). To the end, first we set

$$\tilde{\iota} = \#(\Lambda) \quad \text{and} \quad |\Lambda|_* = 1 + \max_{\lambda \in \Lambda} \check{L}(\lambda) , \quad (3.8)$$

where $\#(\Lambda)$ is the cardinal number of Λ and $\check{L}(\lambda) = \sum_{j=1}^n \lambda(j)$. Now, to choose a weight sequence λ in the set Λ we use the empirical quadratic risk, defined as

$$\text{Err}_n(\lambda) = \| \hat{S}_\lambda - S \|^2,$$

which in our case is equal to

$$\text{Err}_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \hat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \hat{\theta}_{j,n} \theta_j + \sum_{j=1}^{\infty} \theta_j^2. \quad (3.9)$$

Since the Fourier coefficients $(\theta_j)_{j \geq 1}$ are unknown, we replace the terms $\hat{\theta}_{j,n} \theta_j$ by

$$\tilde{\theta}_{j,n} = \hat{\theta}_{j,n}^2 - \frac{\hat{\sigma}_n}{n}, \quad (3.10)$$

where $\hat{\sigma}_n$ is an estimate for the variance proxy σ_Q defined in (2.9). If it is known, we take $\hat{\sigma}_n = \sigma_Q$; otherwise, we can choose it, for example, as in [14], i.e.

$$\hat{\sigma}_n = \sum_{j=[\sqrt{n}]+1}^n \hat{t}_{j,n}^2, \quad (3.11)$$

where $\hat{t}_{j,n}$ are the estimators for the Fourier coefficients with respect to the trigonometric basis (3.2), i.e.

$$\hat{t}_{j,n} = \frac{1}{n} \int_0^n T r_j(t) dy_t. \quad (3.12)$$

Finally, in order to choose the weights, we will minimize the following cost function

$$J_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \hat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \tilde{\theta}_{j,n} + \delta P_n(\lambda), \quad (3.13)$$

where $\delta > 0$ is some threshold which will be specified later and the penalty term is

$$P_n(\lambda) = \frac{\hat{\sigma}_n |\lambda|^2}{n}. \quad (3.14)$$

We define the model selection procedure as

$$\hat{S}_* = \hat{S}_{\hat{\lambda}}, \quad (3.15)$$

where

$$\hat{\lambda} = \operatorname{argmin}_{\lambda \in \Lambda} J_n(\lambda). \quad (3.16)$$

We recall that the set Λ is finite so $\hat{\lambda}$ exists. In the case when $\hat{\lambda}$ is not unique, we take one of them.

Let us now specify the weight coefficients $(\lambda(j))_{1 \leq j \leq n}$. Consider, for some fixed $0 < \varepsilon < 1$, a numerical grid of the form

$$\mathcal{A} = \{1, \dots, k^*\} \times \{\varepsilon, \dots, m\varepsilon\}, \quad (3.17)$$

where $m = \lceil 1/\varepsilon^2 \rceil$. We assume that both parameters $k^* \geq 1$ and ε are functions of n , i.e. $k^* = k^*(n)$ and $\varepsilon = \varepsilon(n)$, such that

$$\begin{cases} \lim_{n \rightarrow \infty} k^*(n) = +\infty, & \lim_{n \rightarrow \infty} \frac{k^*(n)}{\ln n} = 0, \\ \lim_{n \rightarrow \infty} \varepsilon(n) = 0 & \text{and} \quad \lim_{n \rightarrow \infty} n^{\check{\delta}} \varepsilon(n) = +\infty \end{cases} \quad (3.18)$$

for any $\check{\delta} > 0$. One can take, for example, for $n \geq 2$

$$\varepsilon(n) = \frac{1}{\ln n} \quad \text{and} \quad k^*(n) = k_0^* + \sqrt{\ln n}, \quad (3.19)$$

where $k_0^* \geq 0$ is some fixed constant and the threshold $\varsigma^*(n)$ is introduced in (2.9). For each $\alpha = (\beta, 1) \in \mathcal{A}$, we introduce the weight sequence

$$\lambda_\alpha = (\lambda_\alpha(j))_{1 \leq j \leq n}$$

with the elements

$$\lambda_\alpha(j) = \mathbf{1}_{\{1 \leq j < j_*\}} + \left(1 - (j/\omega_\alpha)^\beta\right) \mathbf{1}_{\{j_* \leq j \leq \omega_\alpha\}}, \quad (3.20)$$

where $j_* = 1 + \lfloor \ln v_n \rfloor$, $\omega_\alpha = (d_\beta \mathbf{l} v_n)^{1/(2\beta+1)}$,

$$d_\beta = \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta} \quad \text{and} \quad v_n = n/\varsigma^*.$$

Now we define the set Λ as

$$\Lambda = \{\lambda_\alpha, \alpha \in \mathcal{A}\}. \quad (3.21)$$

It will be noted that in this case the cardinal of the set Λ is

$$\check{\iota} = k^* m. \quad (3.22)$$

Moreover, taking into account that $d_\beta < 1$ for $\beta \geq 1$ we obtain for the set (3.21)

$$|\Lambda|_* \leq 1 + \sup_{\alpha \in \mathcal{A}} \omega_\alpha \leq 1 + (v_n/\varepsilon)^{1/3}. \quad (3.23)$$

Remark 3.1. Note that the form (3.20) for the weight coefficients in (3.7) was proposed by Pinsker in [24] for the efficient estimation in the nonadaptive case, i.e. when the regularity parameters of the function S are known. In the adaptive case these weight coefficients are used in [14, 15] to show the asymptotic efficiency for model selection procedures.

4 Main results

In this section we obtain in Theorem 4.3 the non-asymptotic oracle inequality for the quadratic risk (1.3) for the model selection procedure (3.15) and in Theorem 4.4 the non-asymptotic oracle inequality for the robust risk (1.4) for the same model selection procedure (3.15), considered with the coefficients (3.20). We give the lower and upper bound for the robust risk in Theorems 4.5 and 4.7, and also the optimal convergence rate in Corollary 4.8.

Before stating the non-asymptotic oracle inequality, let us first introduce the following parameters which will be used for describing the rest term in the oracle inequalities. For the renewal density ρ defined in (2.7) we set

$$\Upsilon(x) = \rho(x) - \frac{1}{\tilde{\tau}} \quad \text{and} \quad \|\Upsilon\|_1 = \int_0^{+\infty} |\Upsilon(x)| dx, \quad (4.1)$$

where $\tilde{\tau} = \mathbf{E} \tau_1$. In Proposition 5.2 we show that $|\rho|_* = \sup_{t \geq 0} |\rho(t)| < \infty$ and $\|\Upsilon\|_1 < \infty$. So, using this, we can introduce the following parameters

$$\Psi_Q = 4\kappa_Q \check{\imath} + \left(5 + \frac{4\check{\imath}}{\sigma_Q}\right) \left(\sigma_Q \tilde{\tau} \phi_{max}^2 \|\Upsilon\|_1 + \phi_{max}^4 (1 + \sigma_Q^2)^3 \check{\imath}\right) \quad (4.2)$$

and

$$\mathbf{c}_Q^* = \sigma_Q + 2\kappa_Q + \sigma_Q \tilde{\tau} \phi_{max}^2 \|\Upsilon\|_1 + \phi_{max}^4 (1 + \sigma_Q^2)^2 \check{\imath}, \quad (4.3)$$

where $\check{\imath} = (4\tilde{\tau}^2 + 8) \|\Upsilon\|_1 + 5 + 13(1 + \tilde{\tau})^2(1 + |\rho|_*^2)(\mathbf{E} Y_1^4) + 4\Pi(x^4)$. First, let us state the non-asymptotic oracle inequality for the quadratic risk (1.3) for the model selection procedure (3.15).

Theorem 4.1. Assume that Conditions $\mathbf{H}_1)$ – $\mathbf{H}_4)$ hold. Then, for any $n \geq 1$ and $0 < \delta < 1/6$, the estimator of S given in (3.15) satisfies the following oracle inequality

$$\mathcal{R}_Q(\hat{S}_*, S) \leq \frac{1 + 3\delta}{1 - 3\delta} \min_{\lambda \in \Lambda} \mathcal{R}_Q(\hat{S}_\lambda, S) + \frac{\Psi_Q + 10|\Lambda|_* \mathbf{E}_S |\hat{\sigma}_n - \sigma_Q|}{n\delta}. \quad (4.4)$$

Now we study the estimate (3.11).

Proposition 4.2. *Assume that Conditions \mathbf{H}_1 – \mathbf{H}_4 hold and that the function $S(\cdot)$ is continuously differentiable. Then, for any $n \geq 2$,*

$$\mathbf{E}_{Q,S}|\hat{\sigma}_n - \sigma_Q| \leq \frac{6\|\dot{S}\|^2 + \mathbf{c}_Q^*}{\sqrt{n}}. \quad (4.5)$$

Theorem 4.1 and Proposition 4.2 implies the following result.

Theorem 4.3. *Assume that Conditions \mathbf{H}_1 – \mathbf{H}_4 hold and that the function S is continuously differentiable. Then, for any $n \geq 1$ and $0 < \delta \leq 1/6$, the procedure (3.15), (3.11) satisfies the following oracle inequality*

$$\mathcal{R}_Q(\hat{S}_*, S) \leq \frac{1 + 3\delta}{1 - 3\delta} \min_{\lambda \in \Lambda} \mathcal{R}_Q(\hat{S}_\lambda, S) + \frac{60\tilde{\Lambda}_n \|\dot{S}\|^2 + \tilde{\Psi}_{Q,n}}{n\delta}, \quad (4.6)$$

where $\tilde{\Psi}_{Q,n} = 10\tilde{\Lambda}_n \mathbf{c}_Q^* + \Psi_Q$ and $\tilde{\Lambda}_n = |\Lambda|_*/\sqrt{n}$.

Remark 4.1. *Note that the coefficient κ_Q can be estimated as $\kappa_Q \leq (1 + \tilde{\tau}|\rho|_*)\sigma_Q$. Therefore, taking into account that $\phi_{max}^4 \geq 1$, the remainder term in (4.6) can be estimated as*

$$\tilde{\Psi}_{Q,n} \leq \mathbf{C}_* \left(1 + \sigma_Q^6 + \frac{1}{\sigma_Q} \right) (1 + \tilde{\Lambda}_n) \tilde{\iota} \phi_{max}^4, \quad (4.7)$$

where $\mathbf{C}_* > 0$ is some constant which is independent of the distribution Q .

Furthermore, let us study the robust risk (1.4) for the procedure (3.15). In this case, the distribution family \mathcal{Q}_n consists in all distributions on the Skorokhod space $\mathcal{D}[0, n]$ of the process (1.2) with the parameters satisfying the conditions (2.9) and (2.10).

Moreover, we assume also that the number of the weight vectors and the upper bound for the basis functions in (3.1) may depend on $n \geq 1$, i.e. $\tilde{\iota} = \tilde{\iota}(n)$ and $\phi_* = \phi_*(n)$, such that for any $\tilde{\epsilon} > 0$

$$\lim_{n \rightarrow \infty} \frac{\tilde{\iota}(n)}{n^{\tilde{\epsilon}}} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{\phi_*(n)}{n^{\tilde{\epsilon}}} = 0. \quad (4.8)$$

The next result presents the non-asymptotic oracle inequality for the robust risk (1.4) for the model selection procedure (3.15), considered with the coefficients (3.20).

Theorem 4.4. Assume that Conditions $\mathbf{H}_1) - \mathbf{H}_4)$ hold and that the unknown function S is continuously differentiable. Then, for the robust risk defined in (1.4) through the distribution family (2.9) – (2.10), the procedure (3.15) with the coefficients (3.20) for any $n \geq 1$ and $0 < \delta < 1/6$, satisfies the following oracle inequality

$$\mathcal{R}^*(\hat{S}_*, S) \leq \frac{1 + 3\delta}{1 - 3\delta} \min_{\lambda \in \Lambda} \mathcal{R}^*(\hat{S}_\lambda, S) + \frac{\mathbf{U}_n^*(S)}{n\delta}, \quad (4.9)$$

where the sequence $\mathbf{U}_n^*(S) > 0$ is such that, under the conditions (2.10), (3.18) and (4.8), for any $r > 0$ and $\check{\delta} > 0$,

$$\lim_{n \rightarrow \infty} \sup_{\|\hat{S}\| \leq r} \frac{\mathbf{U}_n^*(S)}{n^{\check{\delta}}} = 0. \quad (4.10)$$

Now we study the asymptotic efficiency for the procedure (3.15) with the coefficients (3.20), with respect to the robust risk (1.4) defined by the distribution family (2.9)–(2.10). To this end, we assume that the unknown function S in the model (1.1) belongs to the Sobolev ball

$$W_r^k = \{f \in \mathcal{C}_{per}^k[0, 1] : \sum_{j=0}^k \|f^{(j)}\|^2 \leq \mathbf{r}\}, \quad (4.11)$$

where $\mathbf{r} > 0$ and $k \geq 1$ are some unknown parameters, $\mathcal{C}_{per}^k[0, 1]$ is the set of k times continuously differentiable functions $f : [0, 1] \rightarrow \mathbb{R}$ such that $f^{(i)}(0) = f^{(i)}(1)$ for all $0 \leq i \leq k$. The function class W_r^k can be written as an ellipsoid in $\mathbf{L}_2[0, 1]$, i.e.,

$$W_r^k = \{f \in \mathcal{C}_{per}^k[0, 1] : \sum_{j=1}^{\infty} a_j \theta_j^2 \leq \mathbf{r}\}, \quad (4.12)$$

where $a_j = \sum_{i=0}^k (2\pi[j/2])^{2i}$ and $\theta_j = \int_0^1 f(v) \text{Tr}_j(v) dv$. We recall that the trigonometric basis $(\text{Tr}_j)_{j \geq 1}$ is defined in (3.2).

Similarly to [14, 15] we will show here that the asymptotic sharp lower bound for the robust risk (1.4) is given by

$$\mathbf{r}_k^* = ((2k + 1)\mathbf{r})^{1/(2k+1)} \left(\frac{k}{(k + 1)\pi} \right)^{2k/(2k+1)}. \quad (4.13)$$

Note that this is the well-known Pinsker constant obtained for the nonadaptive filtration problem in “signal + small white noise” model (see, for example, [24]). Let Π_n be the set of all estimators \hat{S}_n measurable with respect to the σ - field $\sigma\{y_t, 0 \leq t \leq n\}$ generated by the process (1.1).

The following two results give the lower and upper bound for the robust risk in our case.

Theorem 4.5. *Under Conditions (2.9) and (2.10),*

$$\liminf_{n \rightarrow \infty} v_n^{2k/(2k+1)} \inf_{\hat{S}_n \in \Pi_n} \sup_{S \in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\hat{S}_n, S) \geq \mathbf{r}_k^*, \quad (4.14)$$

where $v_n = n/\zeta^*$.

Note that if the parameters \mathbf{r} and k are known, i.e. for the non adaptive estimation case, then to obtain the efficient estimation for the "signal+white noise" model Pinsker in [24] proposed to use the estimate \hat{S}_{λ_0} defined in (3.7) with the weights (3.20) in which

$$\lambda_0 = \lambda_{\alpha_0} \quad \text{and} \quad \alpha_0 = (k, \mathbf{l}_0), \quad (4.15)$$

where $\mathbf{l}_0 = \lceil \mathbf{r}/\varepsilon \rceil \varepsilon$. For the model (1.1) – (1.2) we show the same result.

Proposition 4.6. *The estimator \hat{S}_{λ_0} satisfies the following asymptotic upper bound*

$$\lim_{n \rightarrow \infty} v_n^{2k/(2k+1)} \sup_{S \in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\hat{S}_{\lambda_0}, S) \leq \mathbf{r}_k^*.$$

For the adaptive estimation we use the model selection procedure (3.15) with the parameter δ defined as a function of n satisfying

$$\lim_n \delta_n = 0 \quad \text{and} \quad \lim_n n^{\check{\delta}} \delta_n = 0 \quad (4.16)$$

for any $\check{\delta} > 0$. For example, we can take $\delta_n = (6 + \ln n)^{-1}$.

Theorem 4.7. *Assume that Conditions \mathbf{H}_1 – \mathbf{H}_4 hold true. Then the robust risk defined in (1.4) through the distribution family (2.9)–(2.10) for the procedure (3.15) based on the trigonometric basis (3.2) with the coefficients (3.20) and the parameter $\delta = \delta_n$ satisfying (4.16) has the following asymptotic upper bound*

$$\limsup_{n \rightarrow \infty} v_n^{2k/(2k+1)} \sup_{S \in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\hat{S}_*, S) \leq \mathbf{r}_k^*. \quad (4.17)$$

Theorem 4.5 and Theorem 4.7 allow us to compute the optimal convergence rate.

Corollary 4.8. *Under the assumptions of Theorem 4.7, we have*

$$\lim_{n \rightarrow \infty} v_n^{2k/(2k+1)} \inf_{\hat{S}_n \in \Pi_n} \sup_{S \in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\hat{S}_n, S) = \mathbf{r}_k^*. \quad (4.18)$$

Remark 4.2. *It is well known that the optimal (minimax) risk convergence rate for the Sobolev ball $W_{\mathbf{r}}^k$ is $n^{2k/(2k+1)}$ (see, for example, [24], [22]). We see here that the efficient robust rate is $v_n^{2k/(2k+1)}$, i.e., if the distribution upper bound $\zeta^* \rightarrow 0$ as $n \rightarrow \infty$, we obtain a faster rate with respect to $n^{2k/(2k+1)}$, and, if $\zeta^* \rightarrow \infty$ as $n \rightarrow \infty$, we obtain a slower rate. In the case when ζ^* is constant, then the robust rate is the same as the classical non robust convergence rate.*

5 Renewal density

This section is concerned with results related to the renewal measure (2.7). We start with the following lemma.

Lemma 5.1. *Let τ be a positive random variable with a density g , such that $\mathbf{E}e^{\beta\tau} < \infty$ for some $\beta > 0$. Then there exists a constant β_1 , $0 < \beta_1 \leq \beta$ for which,*

$$\mathbf{E}e^{(\beta_1+i\omega)\tau} \neq 1 \quad \forall \omega \in \mathbb{R}.$$

Proof. We will show this lemma by the contradiction, i.e. assume there exists some sequence of positive numbers going to zero $(\gamma_k)_{k \geq 1}$ and a sequence $(w_k)_{k \geq 1}$ such that

$$\mathbf{E}e^{(\gamma_k+iw_k)\tau} = 1 \tag{5.1}$$

for any $k \geq 1$. Firstly assume that $\limsup_{k \rightarrow \infty} w_k = +\infty$. Note that in this case, for any $N \geq 1$,

$$\begin{aligned} \left| \int_0^N e^{\gamma_k t} \cos(w_k t) g(t) dt \right| &\leq \left| \int_0^N \cos(w_k t) g(t) dt \right| \\ &\quad + \left| \int_0^N (e^{\gamma_k t} - 1) \cos(w_k t) g(t) dt \right|, \end{aligned}$$

i.e., in view of Lemma A.4, for any fixed $N \geq 1$

$$\limsup_{k \rightarrow \infty} \int_0^N e^{\gamma_k t} \cos(w_k t) g(t) dt = 0.$$

Since for some $\beta > 0$ the integral $\int_0^{+\infty} e^{\beta t} g(t) dt < \infty$, we get

$$\lim_{k \rightarrow \infty} \int_0^{+\infty} e^{\gamma_k t} \cos(w_k t) g(t) dt = 0.$$

Let now $\limsup_{k \rightarrow \infty} w_k = \omega_\infty \neq 0$ and $0 < |\omega_\infty| < \infty$. In this case there exists a sequence $(l_k)_{k \geq 1}$ such that $\lim_{k \rightarrow \infty} w_{l_k} = \omega_\infty$, i.e.

$$1 = \limsup_{k \rightarrow \infty} \mathbf{E}e^{\gamma_{l_k} \tau} \cos(\tau w_{l_k}) = \mathbf{E} \cos(\tau \omega_\infty).$$

It is clear that, for random variables having density, the last equality is possible if and only if $w_\infty = 0$. In this case, i.e. when $\limsup_{k \rightarrow \infty} w_{l_k} = 0$, the equation (5.1) implies

$$\limsup_{k \rightarrow \infty} \mathbf{E} e^{\gamma_{l_k} \tau} \frac{\sin(\tau w_{l_k})}{w_{l_k}} = \mathbf{E} \tau = 0.$$

But, under our conditions, $\mathbf{E}\tau > 0$. These contradictions imply the desired result.

□

Proposition 5.2. *Let τ be a positive random variable with the distribution η having a density g which satisfies Conditions \mathbf{H}_1)– \mathbf{H}_4). Then the renewal measure (2.7) is absolutely continuous with density ρ , for which*

$$\rho(x) = \frac{1}{\tilde{\tau}} + \Upsilon(x), \quad (5.2)$$

where $\tilde{\tau} = \mathbf{E}\tau_1$ and $\Upsilon(\cdot)$ is some function defined on \mathbb{R}_+ with values in \mathbb{R} such that

$$\sup_{x \geq 0} x^\gamma |\Upsilon(x)| < \infty \quad \text{for all } \gamma > 0.$$

Proof. First note, that we can represent the renewal measure $\check{\eta}$ as $\check{\eta} = \eta * \eta_0$ and $\eta_0 = \sum_{j=0}^{\infty} \eta^{(j)}$. It is clear that in this case the density ρ of $\check{\eta}$ can be written as

$$\rho(x) = \int_0^x g(x-y) \sum_{n \geq 0} g^{(n)}(y) dy. \quad (5.3)$$

Now we use the arguments proposed in the proof of Lemma 9.5 from [5]. For any $0 < \epsilon < 1$ we set

$$\rho_\epsilon(x) = \int_0^x g(x-y) \left(\sum_{n \geq 0} (1-\epsilon)^n g^{(n)}(y) - \frac{(1-\epsilon)}{\tilde{\tau}} g_0(y) \right) dy - g(x), \quad (5.4)$$

where $g_0(y) = e^{-\epsilon y / \tilde{\tau}} 1_{\{y > 0\}}$. It is easy to deduce that for any $x \in \mathbb{R}$

$$\lim_{\epsilon \rightarrow 0} \rho_\epsilon(x) = \rho(x) - \frac{1}{\tilde{\tau}} \int_0^x g(z) dz - g(x). \quad (5.5)$$

Moreover, in view of the condition \mathbf{H}_1) we obtain that the function $\rho_\epsilon(x)$ satisfies the condition \mathbf{D}) from Section A.2. So, through Proposition A.5 we get

$$\rho_\epsilon(x+) + \rho_\epsilon(x-) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta} \hat{\rho}_\epsilon(\theta) d\theta,$$

where $\hat{\rho}_\epsilon(\theta) = \int_{\mathbb{R}} e^{i\theta x} \rho_\epsilon(x) dx$. Note that

$$|\hat{g}(\theta)| = \left| \int_{\mathbb{R}} e^{i\theta x} g(x) dx \right| \leq \int_{\mathbb{R}} g(x) dx = 1,$$

i.e. for any $0 < \epsilon < 1$ we have $|1 - (1 - \epsilon)\widehat{g}(\theta)| < 1$ and therefore

$$\sum_{n=0}^{\infty} (1 - \epsilon)^n (\widehat{g}(\theta))^n = \frac{1}{1 - (1 - \epsilon)\widehat{g}(\theta)}.$$

From this and, taking into account that

$$\widehat{g}_0(\theta) = \int_{\mathbb{R}} e^{i\theta x} g_0(x) dx = \frac{\check{\tau}}{\epsilon - i\check{\tau}\theta},$$

we obtain

$$\begin{aligned} \widehat{\rho}_\epsilon(\theta) &= \widehat{g}(\theta) \sum_{n=0}^{\infty} (1 - \epsilon)^n (\widehat{g}(\theta))^n - \left(\frac{1 - \epsilon}{\check{\tau}} \right) \widehat{g}(\theta) \widehat{g}_0(\theta) - \widehat{g}(\theta) \\ &= \widehat{g}(\theta) G_\epsilon(\theta) \quad \text{and} \quad G_\epsilon(\theta) = \frac{1}{1 - (1 - \epsilon)\widehat{g}(\theta)} - \frac{1 - i\check{\tau}\theta}{\epsilon - i\check{\tau}\theta}, \end{aligned}$$

i.e.

$$\rho_\epsilon(x-) + \rho_\epsilon(x+) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta} \widehat{g}(\theta) G_\epsilon(\theta) d\theta. \quad (5.6)$$

One can check directly that

$$\sup_{0 < \epsilon < 1, \theta \in \mathbb{R}} |G_\epsilon(\theta)| < \infty.$$

Therefore, using the condition \mathbf{H}_3) and the Lebesgue's dominated convergence theorem, we can pass to limit as $\epsilon \rightarrow 0$ in (5.6), i.e., we obtain that

$$\rho(x+) + \rho(x-) - \frac{2}{\check{\tau}} \int_0^x g(z) dz - g(x+) - g(x-) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta} \widehat{g}(\theta) G_0(\theta) d\theta,$$

where

$$G_0(\theta) = \frac{1}{1 - \widehat{g}(\theta)} + \frac{1 - i\check{\tau}\theta}{i\check{\tau}\theta}.$$

Using here again Proposition A.5 we deduce that

$$\rho(x+) + \rho(x-) = \frac{2}{\check{\tau}} \int_0^x g(z) dz + \frac{1}{\pi} \int_{\mathbb{R}} e^{-ix\theta} \widehat{g}(\theta) \check{G}(\theta) d\theta \quad (5.7)$$

and

$$\check{G}(\theta) = \frac{1}{1 - \widehat{g}(\theta)} + \frac{1}{i\check{\tau}\theta}.$$

Note now that we can represent the density (5.3) as

$$\rho(x) = g * \sum_{n \geq 0} g^{(n)} = \sum_{n \geq 1} g^{(n)}(x) = g(x) + \sum_{n \geq 2} g^{(n)}(x) =: g(x) + \rho_c(x)$$

and the function $\rho_c(x)$ is continuous for all $x \in \mathbb{R}$. This means that

$$\tilde{\rho}(x) = \frac{\rho(x+) + \rho(x-)}{2} - \rho(x) = \frac{g(x+) + g(x-)}{2} - g(x)$$

and, therefore, the condition \mathbf{H}_2) implies that, for any $\gamma > 0$,

$$\sup_{x \geq 0} x^\gamma |\tilde{\rho}(x)| < \infty.$$

Now we can rewrite (5.7) as

$$\rho(x) = \frac{1}{\tau} \int_0^x g(z) dz + \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ix\theta} \hat{g}(\theta) \check{G}(\theta) d\theta - \tilde{\rho}(x). \quad (5.8)$$

Taking into account that $\mathbf{E} e^{\beta\tau} < \infty$ for some $\beta > 0$ we can obtain that

$$\sup_{x \geq 0} x^\gamma \int_x^{+\infty} g(z) dz < \infty.$$

To study the second term in (5.8) we will use Proposition A.3. Indeed, the condition \mathbf{H}_3) implies the first limit equality in (A.1). The second one follows directly from Lemma A.4. Therefore, in view of Proposition A.3, there exists some $\beta^* > 0$ such that, for any $0 \leq \beta_0 \leq \beta^*$,

$$\int_{\mathbb{R}} e^{-ix\theta} \hat{g}(\theta) \check{G}(\theta) d\theta = e^{-\beta_0 x} \int_{\mathbb{R}} e^{-ix\theta} \hat{g}(\theta - i\beta_0) \check{G}(\theta - i\beta_0) d\theta.$$

Note that, due to Lemma 5.1, the function $1 - \hat{g}(\theta)$ has no zeros on the line $\{z \in \mathbb{C} : \text{Im}(z) = -\beta_1\}$. Moreover, one can check directly that $\theta = 0$ is an isolated zero. So, this means that for any $N > 1$ there can be only finitely many zeros in $\{z \in \mathbb{C} : -\beta_1 < \text{Im}(z) < 0, |\text{Re}(z)| < N\}$ of the function $1 - \hat{g}(\theta)$. Moreover, note that in view of lemma A.4 for any $r > 0$

$$\lim_{\text{Re}(\theta) \rightarrow \infty, |\text{Im}(\theta)| \leq r} \hat{g}(\theta) = 0.$$

This means that there exists $N > 0$ such that the function $1 - \hat{g}(\theta) \neq 0$ for $\theta \in \{z \in \mathbb{C} : -\beta_1 < \text{Im}(z) < 0, |\text{Re}(z)| \geq N\}$. So, there can be only finitely many zeros of the function $1 - \hat{g}(\theta)$ in $\{z \in \mathbb{C} : -\beta_1 < \text{Im}(z) < 0\}$ for some

fixed $0 < \beta_1 < \beta$. Therefore, there exists some $\beta_0 > 0$ for which the function $1 - \widehat{g}(\theta)$ has no zeros in $\{z \in \mathbb{C} : -\beta_0 < \text{Im}(z) < 0\}$, i.e. the function $\check{G}(\theta)$ will be bounded in this set and we obtain that

$$\sup_{x \geq 0} e^{\beta_0 x} \left| \int_{\mathbb{R}} e^{-ix\theta} \widehat{g}(\theta) \check{G}(\theta) d\theta \right| < \infty.$$

This the conclusion follows. \square

Using this proposition we can study the renewal process $(N_t)_{t \geq 0}$ introduced in (2.4).

Corollary 5.3. *Assume that Conditions \mathbf{H}_1)– \mathbf{H}_4) hold true. Then, for any $t > 0$,*

$$\mathbf{E} N_t \leq |\rho|_* t \quad \text{and} \quad \mathbf{E} N_t^2 \leq |\rho|_* t + |\rho|_*^2 t^2. \quad (5.9)$$

Proof. First, by means of Proposition 5.2, note that we get

$$\mathbf{E} N_t = \mathbf{E} \sum_{k \geq 1} \mathbf{1}_{\{T_k \leq t\}} = \int_0^t \rho(v) dv \leq |\rho|_* t.$$

Regarding the last bound in (5.9), we use the same reasoning as in the previous inequality, i.e., we obtain

$$\begin{aligned} \mathbf{E} N_t^2 &= \mathbf{E} \sum_{k \geq 1} \mathbf{1}_{\{T_k \leq t\}} + 2\mathbf{E} \sum_{k \geq 1} \mathbf{1}_{\{T_k \leq t\}} \sum_{j=k+1} \mathbf{1}_{\{T_j \leq t\}} \\ &= \mathbf{E} N_t + 2\mathbf{E} \sum_{k \geq 1} \mathbf{1}_{\{T_k \leq t\}} \Theta(T_k) = \mathbf{E} N_t + \int_0^t \Theta(v) \rho(v) dv, \end{aligned}$$

where, for $0 \leq v \leq t$, we defined the function $\Theta(v) = \mathbf{E} N_{t-v} \leq |\rho|_*(t-v)$. \square

6 Stochastic calculus for semi-Markov processes

In this section we give some results of stochastic calculus for the process $(\xi_t)_{t \geq 0}$ given in (1.2), needed all along this paper. As the process ξ_t is the combination of a Levy process and a semi-Markov process, these results are not standard and need to be provided.

Lemma 6.1. *Let f and g be any non-random functions from $\mathbf{L}_2[0, n]$ and $(I_t(f))_{t \geq 0}$ be the process defined in (2.5). Then, for any $0 \leq t \leq n$,*

$$\mathbf{E} I_t(f) I_t(g) = \varrho_1^2(f, g)_t + \varrho_2^2(f, g\rho)_t, \quad (6.1)$$

where $(f, g)_t = \int_0^t f(s) g(s) ds$ and ρ is the density defined in (2.7).

Proof. First, note that we can represent the stochastic integral $I_t(f)$ as

$$I_t(f) = \varrho_1 I_t^L(f) + \varrho_2 I_t^z(f), \quad (6.2)$$

where

$$I_t^L(f) = \int_0^t f(s) dL_s \quad \text{and} \quad I_t^z(f) = \int_0^t f(s) dz_s.$$

Note that the mutual covariation for the martingales $I_t^L(f)$ and $I_t^L(g)$ (see, for example, [18]) may be calculated as

$$[I^L(f), I^L(g)]_t = \check{\varrho}^2 \int_0^t f(s)g(s)ds + (1 - \check{\varrho}^2) \sum_{0 \leq s \leq t} f(s)g(s) (\Delta \check{L}_s)^2, \quad (6.3)$$

where $\Delta \check{L}_s = \check{L}_s - \check{L}_{s-}$. Taking into account that $\mathbf{E} I_t^L(f) I_t^L(g) = \mathbf{E} [I^L(f), I^L(g)]_t$ and that in view of the first condition in (2.2) $\Pi(x^2) = 1$, we obtain that

$$\begin{aligned} \mathbf{E} I_t^L(f) I_t^L(g) &= \check{\varrho}^2 \int_0^t f(s)g(s)ds + (1 - \check{\varrho}^2) \Pi(x^2) \int_0^t f(s)g(s)ds \\ &= \int_0^t f(s)g(s)ds. \end{aligned} \quad (6.4)$$

Moreover, note that

$$\begin{aligned} \mathbf{E} I_t^z(f) I_t^z(g) &= \mathbf{E} \left(\sum_{l=1}^{\infty} f(T_l)g(T_l) Y_l^2 1_{\{T_l \leq t\}} \right) \\ &= \mathbf{E} \left(\sum_{l=1}^{\infty} f(T_l)g(T_l) 1_{\{T_l \leq t\}} \right) = \int_0^t f(s)g(s)\rho(s)ds. \end{aligned}$$

Hence the conclusion follows. \square

Lemma 6.2. Assume that Conditions \mathbf{H}_1 – \mathbf{H}_4 hold true. Then, for any $n \geq 1$ and for any non random function f from $\mathbf{L}_2[0, n]$, the stochastic integral (2.5) exists and satisfies the properties (2.6) with the coefficient \varkappa_Q given in (2.6).

Proof. This lemma follows directly from Lemma 6.1 with $f = g$ and Proposition 5.2. \square

Lemma 6.3. Let f and g be bounded functions defined on $[0, \infty) \times \mathbb{R}$. Then, for any $k \geq 1$,

$$\mathbf{E} \left(I_{T_{k-}}(f) I_{T_{k-}}(g) \mid \mathcal{G} \right) = \varrho_1^2(f, g)_{T_k} + \varrho_2^2 \sum_{l=1}^{k-1} f(T_l) g(T_l),$$

where \mathcal{G} is the σ -field generated by the sequence $(T_l)_{l \geq 1}$, i.e., $\mathcal{G} = \sigma\{T_l, l \geq 1\}$.

Proof. Using (6.2), (6.4) and, taking into account that the process $(L_t)_{t \geq 0}$ is independent of the \mathcal{G} , we obtain

$$\mathbf{E} \left(I_{T_k-}(f) I_{T_k-}(g) \mid \mathcal{G} \right) = \varrho_1^2(f, g)_{T_k} + \mathbf{E} \left(I_{T_k-}^z(f) I_{T_k-}^z(g) \mid \mathcal{G} \right).$$

Moreover,

$$\begin{aligned} \mathbf{E} \left(I_{T_k-}^z(f) I_{T_k-}^z(g) \mid \mathcal{G} \right) &= \mathbf{E} \left(\left(\sum_{l=1}^{k-1} f(T_l) Y_l \right) \left(\sum_{l=1}^{k-1} g(T_l) Y_l \right) \mid \mathcal{G} \right) \\ &= \sum_{l=1}^{k-1} f(T_l) g(T_l). \end{aligned}$$

This we obtain the desired result. \square

Lemma 6.4. Assume that Conditions $\mathbf{H}_1)$ – $\mathbf{H}_4)$ hold true. Then, for any measurable bounded non-random functions f and g , we have

$$\left| \mathbf{E} \int_0^n I_{t-}^2(f) g(t) dm_t \right| \leq 2\varrho_2^2 |g|_* |f|_*^2 \|\Upsilon\|_1 n.$$

Proof. Using the definition of the process $(m_t)_{t \geq 0}$ we can represent this integral as

$$\begin{aligned} \int_0^n I_{t-}^2(f) g(t) dm_t &= \sum_{k \geq 1} I_{T_k-}^2(f) g(T_k) Y_k^2 \mathbf{1}_{\{T_k \leq n\}} \\ &\quad - \int_0^n I_t^2(f) g(t) \rho(t) dt =: V_n - U_n. \end{aligned} \quad (6.5)$$

Note now that

$$\mathbf{E} V_n = \mathbf{E} \sum_{k \geq 1} g(T_k) \mathbf{E} \left(I_{T_k-}^2(f) \mid \mathcal{G} \right) \mathbf{1}_{\{T_k \leq n\}}.$$

Now, using Lemma 6.3 we can represent the last expectation as

$$\mathbf{E} V_n = \varrho_1^2 \mathbf{E} V'_n + \varrho_2^2 \mathbf{E} V''_n, \quad (6.6)$$

where

$$V'_n = \sum_{k \geq 1} g(T_k) \|f\|_{T_k}^2 \mathbf{1}_{\{T_k \leq n\}} \quad \text{and} \quad V''_n = \sum_{k \geq 2} g(T_k) \mathbf{1}_{\{T_k \leq n\}} \sum_{l=1}^{k-1} f^2(T_l).$$

The first term in (6.6) can be represented as

$$\mathbf{E} V_n' = \int_0^n g(t) \|f\|_t^2 \rho(t) dt.$$

To estimate the last expectation in (6.6), note that

$$\mathbf{E} V_n'' = \mathbf{E} \sum_{l \geq 1} f^2(T_l) \bar{g}(T_l) \mathbf{1}_{\{T_l \leq n\}} = \int_0^n f^2(v) \bar{g}(v) \rho(v) dv,$$

where

$$\bar{g}(v) = \mathbf{E} \sum_{k \geq 1} g(v + T_k) \mathbf{1}_{\{T_k \leq n-v\}} = \int_v^n g(t) \rho(t-v) dt.$$

Moreover, using now the representation (6.1), we calculate the expectation of the last term in (6.5)

$$\mathbf{E} U_n = \varrho_1^2 \int_0^n \|f\|_t^2 g(t) \rho(t) dt + \varrho_2^2 \int_0^n \check{f}(t) g(t) \rho(t) dt,$$

where $\check{f}(t) = \int_0^t f^2(s) \rho(s) ds$. This implies that

$$\mathbf{E} \int_0^n I_{t-}^2(f) g(t) dm_t = \varrho_2^2 \int_0^n g(t) \delta(t) dt,$$

where $\delta(t) = \int_0^t f^2(v) (\rho(t-v) - \rho(t)) \rho(v) dv$. Note that, in view of Proposition 5.2, the function δ can be estimated as

$$|\delta(t)| \leq |f|_*^2 |\rho|_* \int_0^t |\Upsilon(t-v) - \Upsilon(t)| dv \leq |f|_*^2 |\rho|_* (\|\Upsilon\|_1 + t|\Upsilon(t)|).$$

Therefore,

$$\left| \mathbf{E} \int_0^n I_{t-}^2(f) g(t) dm_t \right| \leq 2\varrho_2^2 |g|_* |f|_*^2 \|\Upsilon\|_1 n$$

and this finishes the proof. \square

Lemma 6.5. *Assume that Conditions \mathbf{H}_1)– \mathbf{H}_4) hold true. Then, for any measurable bounded non-random functions f and g , one has*

$$\mathbf{E} \int_0^n I_{t-}^2(f) I_{t-}(g) g(t) d\xi_t = 0.$$

Proof. First, note that

$$\int_0^n I_{t-}^2(f) I_{t-}(g) g(t) d\xi_t = \varrho_1 \int_0^n I_t^2(f) I_t(g) g(t) dL_t + \varrho_2 \int_0^n I_{t-}^2(f) I_{t-}(g) g(t) dz_t.$$

Second, we will show that

$$\mathbf{E} \int_0^n I_{t-}^2(f) I_{t-}(g) g(t) dL_t = 0. \quad (6.7)$$

Using the notations (6.2), we set

$$J_1 = \int_0^n I_t^2(f) I_t^L(g) g(t) dL_t \quad \text{and} \quad J_2 = \int_0^n I_t^2(f) I_t^z(g) g(t) dL_t,$$

we obtain that

$$\int_0^n I_t^2(f) I_t(g) g(t) dL_t = \varrho_1 J_1 + \varrho_2 J_2. \quad (6.8)$$

Now let us recall the Novikov inequalities, [23], also referred to as the Bichteler–Jacod inequalities (see [4, 21]) providing bound moments of supremum of purely discontinuous local martingales for any predictable function h and any $p \geq 2$

$$\mathbf{E} \sup_{0 \leq t \leq n} \left| \int_{[0,t] \times \mathbb{R}} h d(\mu - \nu) \right|^p \leq C_p^* \mathbf{E} \check{J}_{p,n}(h), \quad (6.9)$$

where C_p^* is some positive constant and

$$\check{J}_{p,n}(h) = \left(\int_{[0,n] \times \mathbb{R}} h^2 d\nu \right)^{p/2} + \int_{[0,n] \times \mathbb{R}} h^p d\nu.$$

By applying this inequality for the non-random function $h = (s, x) = g(s)x$, and, recalling that $\Pi(x^8) < \infty$, we obtain,

$$\sup_{0 \leq t \leq n} \mathbf{E} \left| I_t^L(g) \right|^8 < \infty.$$

Taking into account that, for any non random square integrated function f , the integral $\left(\int_0^t f(s) dw_s \right)$ is Gaussian with the parameters $\left(0, \int_0^t f^2(s) ds \right)$, we obtain

$$\sup_{0 \leq t \leq n} \mathbf{E} \left| I_t^L(g) \right|^8 < \infty.$$

Finally, by using the Cauchy inequality, we can estimate for any $0 < t \leq$ the following expectation as

$$\mathbf{E} (I_t^L(f))^4 (I_t^L(g))^2 < \sqrt{\mathbf{E} (I_t^L(f))^8} \sqrt{\mathbf{E} (I_t^L(f))^4}$$

i.e.,

$$\sup_{0 \leq t \leq n} \mathbf{E} (I_t^L(f))^4 (I_t^L(g))^2 < \infty .$$

Moreover, taking into account that the processes $(L_t)_{t \geq 0}$ and $(z_t)_{t \geq 0}$ are independent, we obtain that

$$\mathbf{E} (I_t^z(f))^4 (I_t^L(g))^2 = \mathbf{E} (I_t^z(f))^4 \mathbf{E} (I_t^L(g))^2 = \Pi(x^2) \int_0^t g^2(s) ds \mathbf{E} (I_t^z(f))^4 .$$

One can check directly here that, for $t > 0$,

$$\mathbf{E} |I_t^z(f)|^4 \leq |f|_*^4 \mathbf{E} Y_1^4 \mathbf{E} N_t^2 .$$

Note that the last bound in Corollary 5.3 yields $\sup_{0 \leq t \leq n} \mathbf{E} (I_t^z(f))^4 < \infty$ and, therefore,

$$\sup_{0 \leq t \leq n} \mathbf{E} (I_t^z(f))^4 (I_t^L(g))^2 < \infty .$$

It follows directly that $\mathbf{E} J_1 = 0$. Now we study the last term in (6.8). To this end, first note that similarly to the previous reasoning we obtain that

$$\mathbf{E} \int_0^n (I_t^L(f))^2 I_t^z(g) g(t) dL_t = 0 \quad \text{and} \quad \mathbf{E} \int_0^n I_t^L(f) I_t^z(f) I_t^z(g) g(t) dL_t = 0 .$$

Therefore, to show (6.7) one needs to show that

$$\mathbf{E} \int_0^n (I_t^z(f))^2 I_t^z(g) g(t) dL_t = 0 . \tag{6.10}$$

To check this note that, for any $0 < t \leq n$ and for any bounded function f ,

$$I_t^z(f) = \sum_{k=1}^{\infty} f(T_k) Y_k \mathbf{1}_{\{T_k \leq t\}} = \sum_{k=1}^{N_n} f(T_k) Y_k \mathbf{1}_{\{T_k \leq t\}} ,$$

i.e.,

$$\int_0^n (I_t^z(f))^2 I_t^z(g) g(t) dL_t = \sum_{k=1}^{N_n} \sum_{l=1}^{N_n} \sum_{j=1}^{N_n} f(T_k) f(T_l) g(T_j) Y_j Y_l Y_k I_{klj} ,$$

where

$$I_{klj} = \int_0^n \mathbf{1}_{\{T_k \leq t\}} \mathbf{1}_{\{T_l \leq t\}} \mathbf{1}_{\{T_j \leq t\}} dL_t.$$

Taking into account that the $(L_t)_{t \geq 0}$ is independent of the field $\mathcal{G}_z = \sigma\{z_t, t \geq 0\}$, we obtain that $\mathbf{E}(I_{klj}|\mathcal{G}_z) = 0$. Therefore,

$$\begin{aligned} & \mathbf{E} \int_0^n (I_t^z(f))^2 I_t^z(g) g(t) dL_t \\ &= \mathbf{E} \sum_{k=1}^{N_n} \sum_{l=1}^{N_n} \sum_{j=1}^{N_n} f(T_k) f(T_l) g(T_j) Y_j Y_l Y_k \mathbf{E}(I_{klj}|\mathcal{G}_z) = 0. \end{aligned}$$

So, we obtain (6.10) and hence the proof is achieved. \square

7 Properties of the regression model (1.1)

In order to prove the oracle inequalities we need to study the conditions introduced in [14] for the general semi-martingale model (1.1). To this end we set for any $x \in \mathbb{R}^n$ the functions

$$B_{1,Q,n}(x) = \sum_{j=1}^n x_j \left(\mathbf{E}_Q \xi_{j,n}^2 - \sigma_Q \right) \quad \text{and} \quad B_{2,Q,n}(x) = \sum_{j=1}^n x_j \tilde{\xi}_{j,n}, \quad (7.1)$$

where σ_Q is defined in (2.9) and $\tilde{\xi}_{j,n} = \xi_{j,n}^2 - \mathbf{E}_Q \xi_{j,n}^2$.

Proposition 7.1. *Assume that Conditions $\mathbf{H}_1)$ – $\mathbf{H}_4)$ hold. Then*

$$\sup_{x \in [-1,1]^n} |B_{1,Q,n}(x)| \leq \mathbf{C}_{1,Q,n}, \quad (7.2)$$

where $\mathbf{C}_{1,Q,n} = \sigma_Q \check{\tau} \phi_{max}^2 \|\Upsilon\|_1$.

Proof. First, note that from (6.2) we have

$$\xi_{j,n} = \frac{\varrho_1}{\sqrt{n}} I_n^L(\phi_j) + \frac{\varrho_2}{\sqrt{n}} I_n^z(\phi_j).$$

So, using (6.4) we can write that

$$\mathbf{E} \xi_{j,n}^2 = \frac{\varrho_1^2}{n} \int_0^n \phi_j^2(t) dt + \frac{\varrho_2^2}{n} \mathbf{E} \sum_{l=1}^{\infty} \phi_j^2(T_l) \mathbf{1}_{\{T_l \leq n\}}. \quad (7.3)$$

Proposition 5.2 implies

$$\begin{aligned} \mathbf{E} \sum_{l=1}^{\infty} \phi_j^2(T_l) 1_{\{T_l \leq n\}} &= \int_0^n \phi_j^2(x) \rho(x) dx \\ &= \frac{1}{\tilde{\tau}} \int_0^n \phi_j^2(x) dx + \int_0^n \phi_j^2(x) \Upsilon(x) dx. \end{aligned}$$

Note that $\int_0^n \phi_j^2(t) dt = n$. So, in view of the condition (3.1), we obtain

$$\left| \mathbf{E} \xi_{j,n}^2 - \sigma_Q \right| = \frac{\varrho_2^2}{n} \left| \int_0^n \phi_j^2(x) \Upsilon(x) dx \right| \leq \frac{\varrho_2^2}{n} \phi_{max}^2 \|\Upsilon\|_1. \quad (7.4)$$

Estimating here ϱ_2^2 by $\sigma_Q \tilde{\tau}$ we obtain the inequality (7.2) and hence the conclusion follows. \square

Proposition 7.2. *Assume that Conditions $\mathbf{H}_1)$ – $\mathbf{H}_4)$ hold. Then*

$$\sup_{|x| \leq 1} \mathbf{E}_Q B_{2,Q,n}^2(x) \leq \mathbf{C}_{2,Q,n}, \quad (7.5)$$

where $\mathbf{C}_{2,Q,n} = \phi_{max}^4 (1 + \sigma_Q^2)^3 \check{\mathbf{I}}$ and $\check{\mathbf{I}}$ is given in (4.3).

Proof. By Ito's formula one gets

$$dI_t^2(f) = 2I_{t-}(f) dI_t(f) + \varrho_1^2 \check{\varrho}^2 f^2(t) dt + \sum_{0 \leq s \leq t} f^2(s) (\Delta \xi_s^d)^2, \quad (7.6)$$

where $\xi_t^d = \varrho_3 \check{L}_t + \varrho_2 z_t$ and $\varrho_3 = \varrho_1 \sqrt{1 - \check{\varrho}^2}$. Taking into account that the processes $(\check{L}_t)_{t \geq 0}$ and $(z_t)_{t \geq 0}$ are independent and the time of jumps T_k defined in (2.4) has a density, we have $\Delta z_s \Delta \check{L}_s = 0$ a.s. for any $s \geq 0$. Therefore, we can rewrite the differential (7.6) as

$$\begin{aligned} dI_t^2(f) &= 2I_{t-}(f) dI_t(f) + \varrho_1^2 \check{\varrho}^2 f^2(t) dt \\ &\quad + \varrho_3^2 d \sum_{0 \leq s \leq t} f^2(s) (\Delta \check{L}_s)^2 + \varrho_2^2 d \sum_{0 \leq s \leq t} f^2(s) (\Delta z_s)^2. \end{aligned} \quad (7.7)$$

From Lemma 6.2 it follows that

$$\mathbf{E} I_t^2(f) = \varrho_1^2 \int_0^t f^2(s) ds + \varrho_2^2 \int_0^t f^2(s) \rho(s) ds.$$

Therefore, putting

$$\tilde{I}_t(f) = I_t^2(f) - \mathbf{E} I_t^2(f), \quad (7.8)$$

we obtain

$$d\tilde{I}_t(f) = 2I_{t-}(f)f(t)d\xi_t + f^2(t)d\tilde{m}_t, \quad \tilde{m}_t = \varrho_3^2\check{m}_t + \varrho_2^2m_t,$$

where $\check{m}_t = \sum_{0 \leq s \leq t} (\Delta \check{L}_s)^2 - t$ and $m_t = \sum_{0 \leq s \leq t} (\Delta z_s)^2 - \int_0^t \rho(s)ds$. For any non-random vector $x = (x_j)_{1 \leq j \leq n}$ with $\sum_{j=1}^n x_j^2 \leq 1$, we set

$$\bar{I}_t(x) = \sum_{j=1}^n x_j \tilde{I}_t(\phi_j). \quad (7.9)$$

Denoting

$$A_t(x) = \sum_{j=1}^n x_j I_t(\phi_j) \phi_j(t) \quad \text{and} \quad B_t(x) = \sum_{j=1}^n x_j \phi_j^2(t), \quad (7.10)$$

we get the following stochastic differential equation for (7.9)

$$d\bar{I}_t(x) = 2A_{t-}(x)d\xi_t + B_t(x)d\tilde{m}_t, \quad \bar{I}_0(x) = 0.$$

Applying the Ito's formula one obtains

$$\begin{aligned} \mathbf{E} \bar{I}_n^2(x) &= 2\mathbf{E} \int_0^n \bar{I}_{t-}(x) d\bar{I}_t(x) + 4\varrho_1^2 \check{\varrho}^2 \mathbf{E} \int_0^n A_t^2(x) dt \\ &\quad + \varrho_3^2 \mathbf{E} \check{D}_n(x) + \varrho_2^2 \mathbf{E} D_n(x), \end{aligned} \quad (7.11)$$

where $\check{D}_n(x) = \sum_{0 \leq t \leq n} (2A_{t-}(x)\Delta \check{L}_t + \varrho_3^2 B_t(x)(\Delta \check{L}_t)^2)^2$ and $D_n(x) = \sum_{k=1}^{+\infty} \left(2A_{T_k-}(x)Y_k + \varrho_2 B_{T_k-}(x)Y_k^2 \right)^2 \mathbf{1}_{\{T_k \leq n\}}$. Let us now show that

$$\left| \mathbf{E} \int_0^n \bar{I}_{t-}(x) d\bar{I}_t(x) \right| \leq 2\varrho_2^4 \phi_{max}^3 \|\Upsilon\|_1 n^2. \quad (7.12)$$

To this end, note that

$$\begin{aligned} \int_0^n \bar{I}_{t-}(x) d\bar{I}_t(x) &= 2 \sum_{1 \leq j, l \leq n} x_j x_l \int_0^n \tilde{I}_{t-}(\phi_j) I_{t-}(\phi_l) \phi_l(t) d\xi_t \\ &\quad + \sum_{j=1}^n x_j \int_0^n \tilde{I}_{t-}(\phi_j) B_t(x) d\tilde{m}_t. \end{aligned}$$

Using here Lemma 6.5, we get $\mathbf{E} \int_0^n \tilde{I}_{t-}(\phi_j) I_{t-}(\phi_i) \phi_i(t) d\xi_t = 0$. Moreover, the process $(\tilde{m}_t)_{t \geq 0}$ is a martingale, i.e. $\mathbf{E} \int_0^n \tilde{I}_{t-}(\phi_j) B_t(x) d\tilde{m}_t = 0$. Therefore,

$$\mathbf{E} \int_0^n \bar{I}_{t-}(x) d\bar{I}_t(x) = \varrho_2^2 \sum_{j=1}^n x_j \mathbf{E} \int_0^n \tilde{I}_{t-}(\phi_j) B_t(x) dm_t.$$

Taking into account here that for any non-random bounded function f

$$\mathbf{E} \int_0^n f(t) dm_t = 0,$$

we obtain $\mathbf{E} \int_0^n \tilde{I}_{t-}(\phi_j) B_t(x) dm_t = \mathbf{E} \int_0^n I_{t-}^2(\phi_j) B_t(x) dm_t$. So, Lemma 6.4 yields

$$\begin{aligned} \left| \mathbf{E} \int_0^n \tilde{I}_{t-}(\phi_j) B_t(x) dm_t \right| &= \left| \sum_{l=1}^n x_l \mathbf{E} \int_0^n I_{t-}^2(\phi_j) \phi_l^2(t) dm_t \right| \\ &\leq 2 \varrho_2^2 \phi_{max}^3 \|\Upsilon\|_1 \sum_{l=1}^n |x_l| n. \end{aligned}$$

Therefore,

$$\begin{aligned} \left| \mathbf{E} \int_0^n \bar{I}_{t-}(x) d\bar{I}_t(x) \right| &\leq 2 \varrho_2^4 \phi_{max}^3 \|\Upsilon\|_1 n \sum_{1 \leq l, j \leq n} |x_l| |x_j| \\ &= 2 \varrho_2^4 \phi_{max}^3 \|\Upsilon\|_1 n \left(\sum_{l=1}^n |x_l| \right)^2. \end{aligned}$$

Taking into account here that $(\sum_{l=1}^n |x_l|)^2 \leq n \sum_{l=1}^n x_l^2 \leq n$, we obtain (7.12). Reminding that $\Pi(x^2) = 1$ we can calculate directly that

$$\mathbf{E} \check{D}_n(x) = 4 \mathbf{E} \int_0^n A_t^2(x) dt + \varrho_3^4 \Pi(x^4) \int_0^n B_t^2(x) dt. \quad (7.13)$$

Note that, thanks to Lemma 6.1, we obtain that

$$\begin{aligned}
\mathbf{E} \int_0^n A_t^2(x) dt &= \sum_{i,j} x_i x_j \int_0^n \phi_i(t) \phi_j(t) \mathbf{E} I_t \phi_i(t) I_t \phi_j(t) dt \\
&= \sum_{i,j} x_i x_j \int_0^n \phi_i(t) \phi_j(t) \int_0^t \phi_i(v) \phi_j(v) (\varrho_1^2 + \varrho_2^2 \rho(v)) dv \\
&= \frac{1}{2} \varrho_1^2 \sum_{i,j} x_i x_j \left(\int_0^n \phi_i(t) \phi_j(t) dt \right)^2 + \varrho_2^2 A_{1,n}(x) \\
&= \frac{n^2}{2} \varrho_1^2 + \varrho_2^2 A_{1,n}(x),
\end{aligned}$$

where $A_{1,n}(x) = \sum_{i,j} x_i x_j \int_0^n \phi_i(t) \phi_j(t) \left(\int_0^t \phi_i(v) \phi_j(v) \rho(v) dv \right) dt$. This term can be estimated through Proposition 5.2 as

$$\begin{aligned}
|A_{1,n}(x)| &= \left| \frac{n^2}{2\tilde{\tau}} + \sum_{i,j} x_i x_j \int_0^n \phi_i(t) \phi_j(t) \left(\int_0^t \phi_i(v) \phi_j(v) \Upsilon(v) dv \right) dt \right| \\
&\leq \frac{n^2}{2\tilde{\tau}} + n \phi_{max}^4 \|\Upsilon\|_1 \sum_{i,j} |x_i| |x_j| \leq \left(\frac{1}{2\tilde{\tau}} + \phi_{max}^4 \|\Upsilon\|_1 \right) n^2.
\end{aligned}$$

So, reminding that $\sigma_Q = \varrho_1^2 + \varrho_2^2/\tilde{\tau}$ and that $\phi_{max} \geq 1$, we obtain that

$$\begin{aligned}
\mathbf{E} \int_0^n A_t^2(x) dt &\leq \left(\frac{\sigma_Q}{2} + \phi_{max}^4 \|\Upsilon\|_1 \right) n^2 \\
&\leq \left(\frac{1}{4} + \|\Upsilon\|_1 \right) \phi_{max}^4 (1 + \sigma_Q^2) n^2. \tag{7.14}
\end{aligned}$$

Taking into account that

$$\sup_{t \geq 0} B_t^2(x) \leq \phi_{max}^4 \left(\sum_{j=1}^n |x_j| \right)^2 \leq \phi_{max}^4 n, \tag{7.15}$$

that $\phi_{max} \geq 1$, and that $\varrho_1^4 \leq \sigma_Q^2$ we estimate the expectation in (7.13) as

$$\mathbf{E} \check{D}_n \leq 4\phi_{max}^4 (1 + \sigma_Q^2) (1 + \|\Upsilon\|_1 + \Pi(x^4)) n^2. \tag{7.16}$$

Moreover, taking into account that the random variable Y_k is independent of $A_{T_k-}(x)$ and of the field $\mathcal{G} = \sigma\{T_j, j \geq 1\}$ and that $\mathbf{E} \left(A_{T_k-}(x) | \mathcal{G} \right) = 0$, we get

$$\begin{aligned} \mathbf{E} \sum_{k=1}^{+\infty} B_{T_k-}(x) A_{T_k-}(x) Y_k^3 1_{\{T_k \leq n\}} &= \sum_{k=1}^{+\infty} \mathbf{E} \mathbf{E} \left(B_{T_k-}(x) A_{T_k-}(x) Y_k^3 1_{\{T_k \leq n\}} | \mathcal{G} \right) \\ &= \mathbf{E} Y_1^3 \mathbf{E} \sum_{k=1}^{+\infty} B_{T_k-}(x) 1_{\{T_k \leq n\}} \mathbf{E} (A_{T_k-}(x) | \mathcal{G}) = 0. \end{aligned}$$

Therefore,

$$\mathbf{E} D_n(x) = \varrho_2^2 \mathbf{E} Y_1^4 D_{1,n}(x) + 4 D_{2,n}(x), \quad (7.17)$$

where

$$D_{1,n}(x) = \sum_{k=1}^{+\infty} \mathbf{E} B_{T_k-}^2(x) 1_{\{T_k \leq n\}} \quad \text{and} \quad D_{2,n}(x) = \sum_{k=1}^{+\infty} \mathbf{E} A_{T_k-}^2(x) 1_{\{T_k \leq n\}}.$$

Using the bound (7.15) we can estimate the term $D_{1,n}$ as $D_{1,n}(x) \leq \phi_{max}^4 n \mathbf{E} N_n$. Using here Corollary 5.3, we obtain

$$D_{1,n}(x) \leq |\rho|_* \phi_{max}^4 n^2. \quad (7.18)$$

Now, to estimate the last term in (7.17), note that the process $A_t(x)$ can be rewritten as

$$A_t(x) = \int_0^t Q_x(t, s) d\xi_s, \quad \text{with} \quad Q_x(t, s) = \sum_{j=1}^n x_j \phi_j(s) \phi_j(t). \quad (7.19)$$

Applying Lemma 6.3 again, we obtain for any $k \geq 1$

$$\mathbf{E} \left(A_{T_k-}^2(x) | \mathcal{G} \right) = \varrho_1^2 \int_0^{T_k} Q_x^2(T_k, s) ds + \varrho_2^2 \sum_{j=1}^{k-1} Q_x^2(T_k, T_j).$$

So, we can represent the last term in (7.17) as

$$D_{2,n} = \varrho_1^2 D_{2,n}^{(1)} + \varrho_2^2 D_{2,n}^{(2)}, \quad (7.20)$$

where

$$D_{2,n}^{(1)} = \sum_{k=1}^{+\infty} \mathbf{E} 1_{\{T_k \leq n\}} \int_0^{T_k} Q_x^2(T_k, s) ds$$

and

$$D_{2,n}^{(2)} = \sum_{k=1}^{+\infty} \mathbf{E} \mathbf{1}_{\{T_k \leq n\}} \sum_{j=1}^{k-1} Q_x^2(T_k, T_j).$$

Thanks to Proposition 5.2 we obtain

$$D_{2,n}^{(1)} = \int_0^n \left(\int_0^t Q_x^2(t, s) ds \right) \rho(t) dt \leq |\rho|_* \int_0^n \int_0^n Q_x^2(t, s) ds dt.$$

In view of the definition of Q_x in (7.19), we can rewrite the last integral as

$$\begin{aligned} \int_0^n Q_x^2(t, s) ds &= \sum_{1 \leq i, j \leq n} x_i x_j \phi_i(t) \phi_j(t) \int_0^n \phi_i(s) \phi_j(s) ds \\ &= \sum_{i=1}^n x_i^2 \phi_i^2(t) \int_0^n \phi_i^2(s) ds = n \sum_{i=1}^n x_i^2 \phi_i^2(t). \end{aligned}$$

Since $\sum_{j=1}^n x_j^2 \leq 1$, we obtain that,

$$\int_0^n Q_x^2(t, s) ds \leq \phi_{max}^2 n \quad \text{and} \quad D_{2,n}^{(1)} \leq \phi_{max}^2 |\rho|_* n^2. \quad (7.21)$$

Let us estimate now the last term in (7.20). First, note that we can represent this term as

$$D_{2,n}^{(2)} = \sum_{k=1}^{+\infty} \mathbf{E} \mathbf{1}_{\{T_k \leq n\}} \sum_{j=1}^{k-1} Q_x^2(T_k, T_j) = \sum_{j=1}^{\infty} \mathbf{1}_{\{T_j \leq n\}} G(T_j) = \int_0^n G(t) \rho(t) dt,$$

where

$$\begin{aligned} G(t) &= \sum_{k=1}^{+\infty} \mathbf{E} \mathbf{1}_{\{T_k \leq n\}} Q_x^2((t + T_k), t) = \int_0^n Q_x^2(t + v, t) \rho(v) dv \\ &= \int_t^{n+t} Q_x^2(u, t) \rho(u - t) du. \end{aligned}$$

It is clear that, for any $0 \leq t \leq n$,

$$\int_t^{n+t} Q_x^2(u, u - t) \rho(u) du \leq |\rho|_* \int_0^{2n} Q_x^2(v, t) dv.$$

In view of the inequality (7.21) we obtain

$$\int_0^{2n} Q_x^2(u, t) \, du = \int_0^{2n} Q_x^2(t, u) \, du \leq 2\phi_{max}^2 n.$$

Therefore,

$$\max_{0 \leq t \leq n} G(t) \leq 2|\rho|_* \phi_{max}^2 n \quad \text{and} \quad D_{2,n}^{(2)} \leq 2|\rho|_*^2 \phi_{max}^2 n^2.$$

So, estimating ϱ_2^2 by $\check{\tau}\sigma_Q$ and taking into account that $\mathbf{E}Y_1^4 \geq 1$, we obtain that we obtain that

$$\mathbf{E} D_n(x) \leq 13(1 + \check{\tau})\phi_{max}^4 \mathbf{E}Y_1^4(1 + |\rho|_*^2) n^2 \sigma_Q.$$

Using all these bound in (7.11) we obtain (7.5) and thus the conclusion follows. \square

Remark 7.1. The properties (7.2) and (7.5) are used to obtain the oracle inequalities given in Section 4 (see, for example, [14]).

8 Simulation

In this section we report the results of a Monte Carlo experiment in order to assess the performance of the proposed model selection procedure (3.15). In (1.1) we chose a 1-periodic function which is defined, for $0 \leq t \leq 1$, as

$$S(t) = t \sin(2\pi t) + t^2(1 - t) \cos(4\pi t). \quad (8.1)$$

We simulate the model

$$dy_t = S(t)dt + d\xi_t,$$

where $\xi_t = 0.5dw_t + 0.5dz_t$.

Here z_t is the semi-Markov process defined in (2.3) with a Gaussian $\mathcal{N}(0, 1)$ sequence $(Y_j)_{j \geq 1}$ and $(\tau_k)_{k \geq 1}$ used in (2.4) taken as $\tau_k \sim \chi_3^2$.

We use the model selection procedure (3.15) with the weights (3.20) in which $k^* = 100 + \sqrt{\ln(n)}$, $t_i = i/\ln(n)$, $m = \lfloor \ln^2(n) \rfloor$ and $\delta = (3 + \ln(n))^{-2}$. We define the empirical risk as

$$\overline{\mathbf{R}} = \frac{1}{p} \sum_{j=1}^p \hat{\mathbf{E}} \left(\hat{S}_n(t_j) - S(t_j) \right)^2, \quad (8.2)$$

where the observation frequency $p = 100001$ and the expectation was taken as an average over $N = 10000$ replications, i.e.,

$$\hat{\mathbf{E}} \left(\hat{S}_n(\cdot) - S(\cdot) \right)^2 = \frac{1}{N} \sum_{l=1}^N \left(\hat{S}_n^l(\cdot) - S(\cdot) \right)^2.$$

n	$\overline{\mathbf{R}}$	$\overline{\mathbf{R}}_*$
20	0.04430	0.235
100	0.01290	0.068
200	0.00812	0.043
1000	0.00196	0.010

Table 1: Empirical risks

We set the relative quadratic risk as

$$\overline{\mathbf{R}}_* = \overline{\mathbf{R}} / \|S\|_p^2, \quad \text{with} \quad \|S\|_p^2 = \frac{1}{p} \sum_{j=0}^p S^2(t_j). \quad (8.3)$$

In our case $\|S\|_p^2 = 0.1883601$.

Table 1 gives the values for the sample risks (8.2) and (8.3) for different numbers of observations n .

Figures 1–4 show the behaviour of the regression function and its estimates by the model selection procedure (3.15) depending on the values of observation periods n . The black full line is the regression function (8.1) and the red dotted line is the associated estimator.

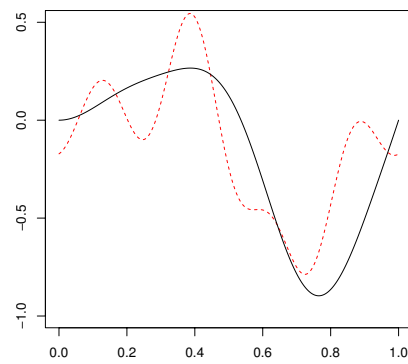


Figure 1: Estimator of S for $n = 20$

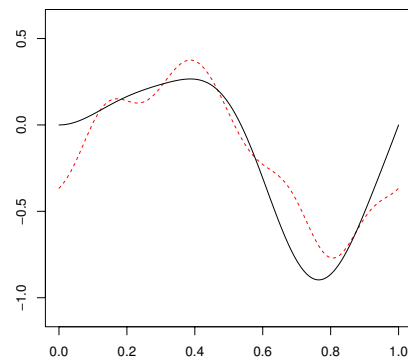


Figure 2: Estimator of S for $n = 100$

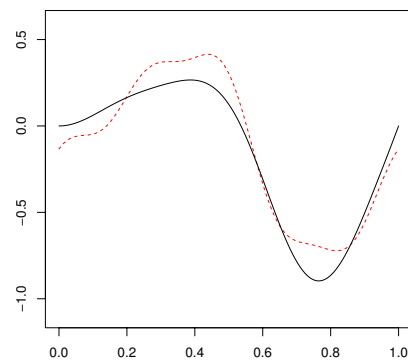


Figure 3: Estimator of S for $n = 200$

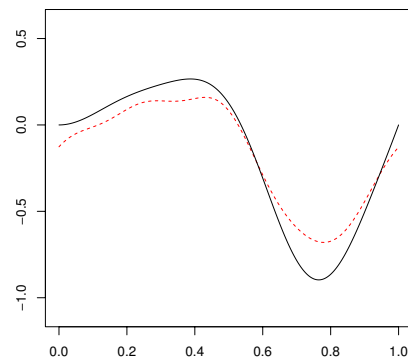


Figure 4: Estimator of S for $n = 1000$

Remark 8.1. From numerical simulations of the procedure (3.15) with various observation numbers n we may conclude that the quality of the proposed procedure: (i) is good for practical needs, i.e. for reasonable (non large) number of observations; (ii) is improving as the number of observations increases.

9 Proofs

We will prove here most of the results of this paper.

9.1 Proof of Theorem 4.1

First, note that we can rewrite the empirical squared error in (3.9) as follows

$$\text{Err}_n(\lambda) = J_n(\lambda) + 2 \sum_{j=1}^{\infty} \lambda(j) \check{\theta}_{j,n} + \|S\|^2 - \delta P_n(\lambda), \quad (9.1)$$

where $\check{\theta}_{j,n} = \tilde{\theta}_{j,n} - \theta_j \hat{\theta}_{j,n}$. Using the definition of $\tilde{\theta}_{j,n}$ in (3.10) we obtain that

$$\check{\theta}_{j,n} = \frac{1}{\sqrt{n}} \theta_j \xi_{j,n} + \frac{1}{n} \tilde{\xi}_{j,n} + \frac{1}{n} \varsigma_{j,n} + \frac{\sigma_Q - \hat{\sigma}_n}{n},$$

where $\varsigma_{j,n} = \mathbf{E}_Q \xi_{j,n}^2 - \sigma_Q$ and $\tilde{\xi}_{j,n} = \xi_{j,n}^2 - \mathbf{E}_Q \xi_{j,n}^2$. Putting

$$M(\lambda) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \lambda(j) \theta_j \xi_{j,n} \quad \text{and} \quad P_n^0 = \frac{\sigma_Q |\lambda|^2}{n}, \quad (9.2)$$

we can rewrite (9.1) as

$$\begin{aligned} \text{Err}_n(\lambda) = & J_n(\lambda) + 2 \frac{\sigma_Q - \hat{\sigma}_n}{n} \check{L}(\lambda) + 2M(\lambda) + \frac{2}{n} B_{1,Q,n}(\lambda) \\ & + 2\sqrt{P_n^0(\lambda)} \frac{B_{2,Q,n}(e(\lambda))}{\sqrt{\sigma_Q n}} + \|S\|^2 - \rho P_n(\lambda), \end{aligned} \quad (9.3)$$

where $e(\lambda) = \lambda/|\lambda|$, the function $\check{L}(\cdot)$ is defined in (3.8) and the functions $B_{1,Q,n}(\cdot)$ and $B_{2,Q,n}(\cdot)$ are given in (7.1).

Let $\lambda_0 = (\lambda_0(j))_{1 \leq j \leq n}$ be a fixed sequence in Λ and $\hat{\lambda}$ be as in (3.16). Substituting λ_0 and $\hat{\lambda}$ in Equation (9.3), we obtain

$$\begin{aligned} \text{Err}_n(\hat{\lambda}) - \text{Err}_n(\lambda_0) &= J(\hat{\lambda}) - J(\lambda_0) + 2 \frac{\sigma_Q - \hat{\sigma}_Q}{n} \check{L}(\varpi) + \frac{2}{n} B_{1,Q,n}(\varpi) + 2M(\varpi) \\ &\quad + 2 \sqrt{P_n^0(\hat{\lambda})} \frac{B_{2,Q,n}(\hat{e})}{\sqrt{\sigma_Q n}} - 2 \sqrt{P_n^0(\lambda_0)} \frac{B_{2,Q,n}(e_0)}{\sqrt{\sigma_Q n}} \\ &\quad - \delta P_n(\hat{\lambda}) + \delta P_n(\lambda_0), \end{aligned} \quad (9.4)$$

where $\varpi = \hat{\lambda} - \lambda_0$, $\hat{e} = e(\hat{\lambda})$ and $e_0 = e(\lambda_0)$. Note that, by (3.8),

$$|\check{L}(\hat{x})| \leq \check{L}(\hat{\lambda}) + \check{L}(\lambda) \leq 2|\Lambda|_*.$$

Applying the inequality

$$2|ab| \leq \delta a^2 + \delta^{-1} b^2 \quad (9.5)$$

implies that, for any $\lambda \in \Lambda$,

$$2 \sqrt{P_n^0(\lambda)} \frac{|B_{2,Q,n}(e(\lambda))|}{\sqrt{\sigma_Q n}} \leq \delta P_n^0(\lambda) + \frac{B_{2,Q,n}^2(e(\lambda))}{\delta \sigma_Q n}.$$

Taking into account the bound (7.2), we get

$$\begin{aligned} \text{Err}_n(\hat{\lambda}) &\leq \text{Err}_n(\lambda_0) + 2M(\varpi) + \frac{2\mathbf{C}_{1,Q,n}}{n} + \frac{2B_{2,Q,n}^*}{\delta \sigma_Q n} \\ &\quad + \frac{1}{n} |\hat{\sigma} - \sigma_Q| (|\hat{\lambda}|^2 + |\lambda_0|^2) + 2\delta P_n(\lambda_0), \end{aligned}$$

where $B_{2,Q,n}^* = \sup_{\lambda \in \Lambda} B_{2,Q,n}^2(e(\lambda))$. Moreover, noting that in view of (3.8) $\sup_{\lambda \in \Lambda} |\lambda|^2 \leq |\Lambda|_*$, we can rewrite the previous bound as

$$\begin{aligned} \text{Err}_n(\hat{\lambda}) &\leq \text{Err}_n(\lambda_0) + 2M(\varpi) + \frac{2\mathbf{C}_{1,Q,n}}{n} + \frac{2B_{2,Q,n}^*}{\delta \sigma_Q n} \\ &\quad + \frac{4|\Lambda|_*}{n} |\hat{\sigma} - \sigma_Q| + 2\delta P_n(\lambda_0). \end{aligned} \quad (9.6)$$

To estimate the second term in the right side of this inequality we set

$$S_x = \sum_{j=1}^n x(j) \theta_j \phi_j, \quad x = (x(j))_{1 \leq j \leq n} \in \mathbb{R}^n.$$

Thanks to (2.6) we estimate the term $M(x)$ for any $x \in \mathbb{R}^n$ as

$$\mathbf{E}_Q M^2(x) \leq \kappa_Q \frac{1}{n} \sum_{j=1}^n x^2(j) \theta_j^2 = \kappa_Q \frac{1}{n} \|S_x\|^2. \quad (9.7)$$

To estimate this function for a random vector $x \in \mathbb{R}^n$ we set

$$Z^* = \sup_{x \in \Lambda_1} \frac{n M^2(x)}{\|S_x\|^2}, \quad \Lambda_1 = \Lambda - \lambda_0.$$

So, through Inequality (9.5), we get

$$2|M(x)| \leq \delta \|S_x\|^2 + \frac{Z^*}{n\delta}. \quad (9.8)$$

It is clear that the last term here can be estimated as

$$\mathbf{E}_Q Z^* \leq \sum_{x \in \Lambda_1} \frac{n \mathbf{E}_Q M^2(x)}{\|S_x\|^2} \leq \sum_{x \in \Lambda_1} \kappa_Q = \kappa_Q \tilde{l}, \quad (9.9)$$

where $\tilde{l} = \text{card}(\Lambda)$. Moreover, note that, for any $x \in \Lambda_1$,

$$\|S_x\|^2 - \|\hat{S}_x\|^2 = \sum_{j=1}^n x^2(j) (\theta_j^2 - \hat{\theta}_j^2) \leq -2M_1(x), \quad (9.10)$$

where $M_1(x) = n^{-1/2} \sum_{j=1}^n x^2(j) \theta_j \xi_{j,n}$. Taking into account that, for any $x \in \Lambda_1$ the components $|x(j)| \leq 1$, we can estimate this term as in (9.7), i.e.,

$$\mathbf{E}_Q M_1^2(x) \leq \kappa_Q \frac{\|S_x\|^2}{n}.$$

Similarly to the previous reasoning we set

$$Z_1^* = \sup_{x \in \Lambda_1} \frac{n M_1^2(x)}{\|S_x\|^2}$$

and we get

$$\mathbf{E}_Q Z_1^* \leq \kappa_Q \tilde{l}. \quad (9.11)$$

Using the same type of arguments as in (9.8), we can derive

$$2|M_1(x)| \leq \delta \|S_x\|^2 + \frac{Z_1^*}{n\delta}. \quad (9.12)$$

From here and (9.10), we get

$$\|S_x\|^2 \leq \frac{\|\widehat{S}_x\|^2}{1-\delta} + \frac{Z_1^*}{n\delta(1-\delta)} \quad (9.13)$$

for any $0 < \delta < 1$. Using this bound in (9.8) yields

$$2M(x) \leq \frac{\delta\|\widehat{S}_x\|^2}{1-\delta} + \frac{Z^* + Z_1^*}{n\delta(1-\delta)}.$$

Taking into account that $\|\widehat{S}_\varpi\|^2 \leq 2(\text{Err}_n(\widehat{\lambda}) + \text{Err}_n(\lambda_0))$, we obtain

$$2M(\varpi) \leq \frac{2\delta(\text{Err}_n(\widehat{\lambda}) + \text{Err}_n(\lambda_0))}{1-\delta} + \frac{Z^* + Z_1^*}{n\delta(1-\delta)}.$$

Using this bound in (9.6) we obtain

$$\begin{aligned} \text{Err}_n(\widehat{\lambda}) &\leq \frac{1+\delta}{1-3\delta} \text{Err}_n(\lambda_0) + \frac{Z^* + Z_1^*}{n\delta(1-3\delta)} + \frac{2\mathbf{C}_{1,Q,n}}{n(1-3\delta)} + \frac{2B_{2,Q,n}^*}{\delta(1-3\delta)\sigma_Q n} \\ &\quad + \frac{(4|\Lambda|_* + 2)}{n(1-3\delta)} |\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)} P_n^0(\lambda_0). \end{aligned}$$

Moreover, for $0 < \delta < 1/6$, we can rewrite this inequality as

$$\begin{aligned} \text{Err}_n(\widehat{\lambda}) &\leq \frac{1+\delta}{1-3\delta} \text{Err}_n(\lambda_0) + \frac{2(Z^* + Z_1^*)}{n\delta} + \frac{4\mathbf{C}_{1,Q,n}}{n} + \frac{4B_{2,Q,n}^*}{\delta\sigma_Q n} \\ &\quad + \frac{(8|\Lambda|_* + 2)}{n} |\widehat{\sigma}_n - \sigma_Q| + \frac{2\delta}{(1-3\delta)} P_n^0(\lambda_0). \end{aligned}$$

In view of Proposition 7.2 we estimate the expectation of the term $B_{2,Q,n}^*$ in (9.6) as

$$\mathbf{E}_Q B_{2,Q,n}^* \leq \sum_{\lambda \in \Lambda} \mathbf{E}_Q B_{2,Q,n}^2(e(\lambda)) \leq i\mathbf{C}_{2,Q,n}.$$

Taking into account that $|\Lambda|_* \geq 1$, we get

$$\begin{aligned} \mathcal{R}(\widehat{S}_*, S) &\leq \frac{1+\delta}{1-3\delta} \mathcal{R}(\widehat{S}_{\lambda_0}, S) + \frac{4\kappa_Q i}{n\delta} + \frac{4\mathbf{C}_{1,Q,n}}{n} + \frac{4i\mathbf{C}_{2,Q,n}}{\delta\sigma_Q n} \\ &\quad + \frac{10|\Lambda|_*}{n} \mathbf{E}_Q |\widehat{\sigma} - \sigma_Q| + \frac{2\delta}{(1-3\delta)} P_n^0(\lambda_0). \end{aligned}$$

Using the upper bound for $P_n(\lambda_0)$ in Lemma A.1, one obtains (4.1), that finishes the proof. \square

9.2 Proof of Proposition 4.2

We use here the same method as in [11]. First of all note that the definition (3.12) implies that

$$\hat{t}_{j,n} = t_j + \frac{1}{\sqrt{n}} \eta_{j,n}, \quad (9.14)$$

where

$$t_j = \int_0^1 S(t) \operatorname{Tr}_j(t) dt \quad \text{and} \quad \eta_{j,n} = \frac{1}{\sqrt{n}} \int_0^n \operatorname{Tr}_j(t) d\xi_t.$$

So, we have

$$\hat{\sigma}_n = \sum_{j=[\sqrt{n}]+1}^n t_j^2 + 2M_n + \frac{1}{n} \sum_{j=[\sqrt{n}]+1}^n \eta_{j,n}^2, \quad (9.15)$$

where

$$M_n = \frac{1}{\sqrt{n}} \sum_{j=[\sqrt{n}]+1}^n t_j \eta_{j,n}.$$

Note that, for continuously differentiable functions (see, for example, Lemma A.6 in [11]), the Fourier coefficients (t_j) satisfy the following inequality, for any $n \geq 1$,

$$\sum_{j=[\sqrt{n}]+1}^{\infty} t_j^2 \leq \frac{4 \left(\int_0^1 |\dot{S}(t)| dt \right)^2}{\sqrt{n}} \leq \frac{4 \|\dot{S}\|^2}{\sqrt{n}}. \quad (9.16)$$

In the same way as in (9.7) we estimate the term M_n , i.e.,

$$\mathbf{E}_Q M_n^2 \leq \frac{\varkappa_Q}{n} \sum_{j=[\sqrt{n}]+1}^n t_j^2 \leq \frac{4 \varkappa_Q \|\dot{S}\|^2}{n \sqrt{n}},$$

while the absolute value of this term for $n \geq 1$ can be estimated as

$$|\mathbf{E}_Q M_n| \leq \frac{\varkappa_Q + \|\dot{S}\|^2}{\sqrt{n}}.$$

Moreover, using Propositions 7.1 and 7.2 we can represent the last term in (9.15) as

$$\frac{1}{n} \sum_{j=[\sqrt{n}]+1}^n \eta_{j,n}^2 = \frac{\sigma_Q(n - \sqrt{n})}{n} + \frac{B_{1,Q,n}(x')}{n} + \frac{B_{2,Q,n}(x'')}{\sqrt{n}},$$

with $x'_j = \mathbf{1}_{\{\sqrt{n} < j \leq n\}}$ and $x''_j = \mathbf{1}_{\{\sqrt{n} < j \leq n\}}/\sqrt{n}$. Therefore,

$$\mathbf{E}_Q \left| \frac{1}{n} \sum_{j=[\sqrt{n}]+1}^n \eta_{j,n}^2 - \sigma_Q \right| \leq \frac{\sigma_Q}{\sqrt{n}} + \frac{\mathbf{C}_{1,Q,n}}{n} + \frac{\sqrt{\mathbf{C}_{2,Q,n}}}{\sqrt{n}}.$$

Taking into account that $\mathbf{C}_{2,Q,n} \geq 1$, we obtain the bound (4.5) and hence the desired result. \square

9.3 Proof of Theorem 4.4

First note, that in view of (3.22) and (3.18)

$$\lim_{n \rightarrow \infty} \frac{\check{l}}{n^{\check{\epsilon}}} = \lim_{n \rightarrow \infty} \frac{k^* m}{n^{\check{\epsilon}}} = 0 \quad \text{for any } \check{\epsilon} > 0.$$

Furthermore, the bound (3.23) and the conditions (2.10) and (3.18) yield

$$\lim_{n \rightarrow \infty} \frac{|\Lambda|_*}{n^{1/3+\check{\epsilon}}} = 0 \quad \text{for any } \check{\epsilon} > 0.$$

So, from here we obtain the convergence (4.10). \square

9.4 Proof of Theorem 4.5

First, we denote by Q_0 the distribution of the noise (1.2) and (2.1) with the parameter $\varrho_1 = \varsigma^*$, $\check{\varrho} = 1$ and $\varrho_2 = 0$, i.e. the distribution for the “signal + white noise” model. So, we can estimate as below the robust risk

$$\mathcal{R}_n^*(\tilde{S}_n, S) \geq \mathcal{R}_{Q_0}(\tilde{S}_n, S).$$

Now Theorem 6.1 from [12] yields the lower bound (4.14). Hence this finishes the proof. \square

9.5 Proof of Proposition 4.6

Putting $\lambda_0(j) = 0$ for $j \geq n$ we can represent the quadratic risk for the estimator (3.7) as

$$\| \hat{S}_{\lambda_0} - S \|^2 = \sum_{j=1}^{\infty} (1 - \lambda_0(j))^2 \theta_j^2 - 2H_n + \frac{1}{n} \sum_{j=1}^n \lambda_0^2(j) \xi_{j,n}^2,$$

where $H_n = n^{-1/2} \sum_{j=1}^n (1 - \lambda_0(j)) \lambda_0(j) \theta_j \xi_{j,n}$. Note that $\mathbf{E}_Q H_n = 0$ for any $Q \in \mathcal{Q}_n$, therefore,

$$\mathbf{E}_Q \|\widehat{S}_{\lambda_0} - S\|^2 = \sum_{j=1}^{\infty} (1 - \lambda_0(j))^2 \theta_j^2 + \frac{1}{n} \mathbf{E}_Q \sum_{j=1}^n \lambda_0^2(j) \xi_{j,n}^2.$$

Proposition 7.1 and the last inequality in (2.9) imply that for any $Q \in \mathcal{Q}_n$

$$\mathbf{E}_Q \sum_{j=1}^n \lambda_0^2(j) \xi_{j,n}^2 \leq \varsigma^* \sum_{j=1}^n \lambda_0^2(j) + \frac{\phi_{max}^2 \varsigma^* \|\Upsilon\|_1}{\check{\tau}} := \varsigma^* \sum_{j=1}^n \lambda_0^2(j) + \mathbf{C}_{1,n}^*.$$

Therefore,

$$\mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \leq \sum_{j=j_*}^{\infty} (1 - \lambda_0(j))^2 \theta_j^2 + \frac{1}{v_n} \sum_{j=1}^n \lambda_0^2(j) + \frac{\mathbf{C}_{1,n}^*}{n},$$

where j_* and v_n are defined in (3.20). Setting

$$\Upsilon_{1,n}(S) = v_n^{2k/(2k+1)} \sum_{j=j_*}^{\infty} (1 - \lambda_0(j))^2 \theta_j^2 \quad \text{and} \quad \Upsilon_{2,n} = \frac{1}{v_n^{1/(2k+1)}} \sum_{j=1}^n \lambda_0^2(j),$$

we rewrite the last inequality as

$$v_n^{2k/(2k+1)} \mathcal{R}_n^*(\widehat{S}_{\lambda_0}, S) \leq \Upsilon_{1,n}(S) + \Upsilon_{2,n} + \check{\mathbf{C}}_n, \quad (9.17)$$

where $\check{\mathbf{C}}_n = v_n^{2k/(2k+1)} \mathbf{C}_{1,n}^*/n$. Note, that the conditions (2.10) and (4.8) imply that $\mathbf{C}_{1,n}^* = o(n^{\check{\delta}})$ as $n \rightarrow \infty$ for any $\check{\delta} > 0$; therefore, $\check{\mathbf{C}}_n \rightarrow 0$ as $n \rightarrow \infty$. Putting

$$u_n = v_n^{2k/(2k+1)} \sup_{j \geq j_*} (1 - \lambda_0(j))^2 / a_j,$$

with a_j defined in (4.12), we estimate the first term in (9.17) as

$$\sup_{S \in W_{\mathbf{r}}^k} \Upsilon_{1,n}(S) \leq \sup_{S \in W_{\mathbf{r}}^k} u_n \sum_{j \geq 1} a_j \theta_j \leq u_n \mathbf{r}.$$

Taking into account that $a_j/(\pi^{2k} j^{2k}) \rightarrow 1$ as $j \rightarrow \infty$ and $\mathbf{l}_0 \rightarrow \mathbf{r}$ as $\varepsilon \rightarrow 0$ and using the definition of ω_{α_0} in (3.20), we obtain that

$$\begin{aligned} \limsup_{n \rightarrow \infty} u_n &\leq \lim_{n \rightarrow \infty} v_n^{2k/(2k+1)} \sup_{j \geq j_*} \frac{(1 - \lambda_0(j))^2}{(\pi j)^{2k}} \\ &= \lim_{n \rightarrow \infty} \frac{v_n^{2k/(2k+1)}}{\pi^{2k} \omega_{\alpha_0}^{2k}} = \frac{1}{\pi^{2k} (\mathbf{d}_k \mathbf{r})^{2k/(2k+1)}}. \end{aligned}$$

Therefore,

$$\limsup_{n \rightarrow \infty} \sup_{S \in W_{\mathbf{r}}^k} \Upsilon_{1,n}(S) \leq \frac{r^{1/(2k+1)}}{\pi^{2k} (d_k)^{2k/(2k+1)}} =: \Upsilon_1^*. \quad (9.18)$$

As to the second term in (9.17), note that

$$\lim_{n \rightarrow \infty} \frac{1}{\omega_{\alpha_0}} \sum_{j=1}^n \lambda_0^2(j) = \int_0^1 (1 - t^k)^2 dt = \frac{2k^2}{(k+1)(2k+1)}.$$

So, taking into account that $\omega_{\alpha_0}/v_n^{1/(2k+1)} \rightarrow (d_k \mathbf{r})^{1/(2k+1)}$ as $n \rightarrow \infty$, the limit of $\Upsilon_{2,n}$ can be calculated as

$$\lim_{n \rightarrow \infty} \Upsilon_{2,n} = \frac{2(d_k \mathbf{r})^{1/(2k+1)} k^2}{(k+1)(2k+1)} =: \Upsilon_2^*.$$

Moreover, since $\Upsilon_1^* + \Upsilon_2^* =: \mathbf{r}_k^*$, we obtain

$$\lim_{n \rightarrow \infty} v_n^{2k/(2k+1)} \sup_{S \in W_{\mathbf{r}}^k} \mathcal{R}_n^*(\hat{S}_{\lambda_0}, S) \leq \mathbf{r}_k^*$$

and get the desired result. \square

9.6 Proof of Theorem 4.7

Combining Proposition 4.6 and Theorem 4.4 yields Theorem 4.6. \square

Acknowledgments. This research was partially supported by the Ministry of Education and Science of the Russian Federation, project (No 2.3208.2017/PCH), by the Russian Federal Professor program of the Ministry of Education and Science of the Russian Federation, (project No 1.472.2016/FPM) and by the Academic D.I. Mendeleev Fund Program of the Tomsk State University (research project NU 8.1.55.2015 L).

10 Appendix

A.1 Property of the penalty term

Lemma A.1. For any $n \geq 1$ and $\lambda \in \Lambda$,

$$P_n^0(\lambda) \leq \mathbf{E}_Q \text{Err}_n(\lambda) + \frac{\mathbf{C}_{1,Q,n}}{n},$$

where the coefficient $P_n^0(\lambda)$ was defined in (9.2).

Proof. By the definition of $\text{Err}_n(\lambda)$ one has

$$\text{Err}_n(\lambda) = \sum_{j=1}^n \left((\lambda(j) - 1)\theta_j + \frac{\lambda(j)}{n} \xi_{j,n} \right)^2.$$

In view of Proposition 7.1, this leads to the desired result

$$\mathbf{E}_Q \text{Err}_n(\lambda) \geq \frac{1}{n} \sum_{j=1}^n \lambda^2(j) \mathbf{E}_Q \xi_{j,n}^2 \geq P_n^0(\gamma) - \frac{\mathbf{C}_{1,Q,n}}{n}.$$

□

A.2 Properties of the Fourier transform

Theorem A.2. *Cauchy (1825)*

Let U be a simply connected open subset of \mathbb{C} , let $g : U \rightarrow \mathbb{C}$ be a holomorphic function, and let γ be a rectifiable path in U whose start point is equal to its end point. Then

$$\oint_{\gamma} g(z) dz = 0.$$

Proposition A.3. Let $g : \mathbb{C} \rightarrow \mathbb{C}$ be a holomorphic function in $U = \{z \in \mathbb{C} : -\beta_1 < \text{Im} z < \beta_2\}$ for some $\beta_1 > 0$ and $\beta_2 > 0$. Assume that, for any $-\beta_1 \leq t \leq 0$,

$$\int_{\mathbb{R}} |g(\theta + it)| d\theta < \infty \quad \text{and} \quad \lim_{|\theta| \rightarrow \infty} g(\theta + it) = 0. \quad (\text{A.1})$$

Then, for any $x \in \mathbb{R}$ and for any $0 < \beta < \beta_1$,

$$\int_{\mathbb{R}} e^{i\theta x} g(\theta) d\theta = e^{-\beta x} \int_{\mathbb{R}} e^{i\theta x} g(\theta - i\beta) d\theta. \quad (\text{A.2})$$

Proof. First note that the conditions of this theorem imply that

$$\int_{\mathbb{R}} e^{i\theta x} g(\theta) d\theta = \lim_{N \rightarrow \infty} \int_{-N}^N e^{i\theta x} g(\theta) d\theta.$$

We fix now $0 < \beta < \beta_1$ and we set for any $N \geq 1$

$$\begin{aligned} \gamma = & \{z \in \mathbb{C} : -N \leq \text{Re} z \leq N, \text{Im} z = 0\} \cup \{z \in \mathbb{C} : -N \leq \text{Im} z \leq N, \text{Re} z = N\} \\ & \cup \{z \in \mathbb{C} : -N \leq \text{Re} z \leq N, \text{Im} z = -\beta\} \cup \{z \in \mathbb{C} : -\beta \leq \text{Im} z \leq 0, \text{Re} z = -N\}. \end{aligned}$$

Now, in view of the Cauchy theorem, we obtain that for any $N \geq 1$

$$\begin{aligned} \oint_{\gamma} e^{izx} g(z) dz &= \int_{-N}^N e^{i\theta x} g(\theta) d\theta + \int_0^{-\beta} e^{i(N+it)x} g(N+it) dt \\ &+ \int_N^{-N} e^{i(-i\beta+\theta)x} g(-i\beta+\theta) d\theta + \int_{-\beta}^0 e^{i(-N+it)x} g(-N+it) dt = 0. \end{aligned} \quad (\text{A.3})$$

The conditions (A.1) provide that

$$\lim_{N \rightarrow \infty} \int_0^{-\beta} e^{i(N+it)x} g(N+it) dt = \lim_{N \rightarrow \infty} \int_{-\beta}^0 e^{i(-N+it)x} g(-N+it) dt = 0.$$

Therefore, letting $N \rightarrow \infty$ in (A.3) we obtain (A.2). Hence we get the desired result. \square

The following technical lemma is also needed in the present paper.

Lemma A.4. *Let $g : [a, b] \rightarrow \mathbb{R}$ be a function from $\mathbf{L}_1[a, b]$. Then, for any fixed $-\infty \leq a < b \leq +\infty$,*

$$\lim_{N \rightarrow \infty} \int_a^b g(x) \sin(Nx) dx = 0 \quad \text{and} \quad \lim_{N \rightarrow \infty} \int_a^b g(x) \cos(Nx) dx = 0. \quad (\text{A.4})$$

Proof. Let first $-\infty < a < b < +\infty$. Assume that g is continuously differentiable, i.e. $g \in \mathbf{C}^1[a, b]$. Then integrating by parts gives us

$$\int_a^b g(x) \sin(Nx) dx = \frac{1}{N} \left(g(b) \sin(Nb) - g(a) \sin(Na) - \int_a^b g'(x) \cos(Nx) dx \right).$$

So, from this we obtain that

$$\left| \int_a^b g(x) \sin(Nx) dx \right| \leq \frac{|g(a)| + |g(b)| + (b-a) \max_{a \leq x \leq b} |g'(x)|}{N}.$$

This implies the first limit in (A.4) for this case. The second one is obtained similarly. Let now g be any absolutely integrated function on $[a, b]$, i.e. $g \in \mathbf{L}_1[a, b]$. In this case there exists a sequence $g_n \in \mathbf{C}^1[a, b]$ such that

$$\lim_{n \rightarrow \infty} \int_a^b |g(x) - g_n(x)| dx = 0.$$

Therefore, taking into account that for any $n \geq 1$

$$\lim_{N \rightarrow \infty} \int_a^b g_n(x) \sin(Nx) dx = 0,$$

we obtain that

$$\limsup_{n \rightarrow \infty} \left| \int_a^b g(x) \sin(Nx) dx \right| \leq \int_a^b |g(x) - g_n(x)| dx.$$

So, letting in this inequality $n \rightarrow \infty$ we obtain the first limit in (A.4) and, similarly, we obtain the second one. Let now $b = +\infty$ and $a = -\infty$. In this case we obtain that for any $-\infty < a < b < +\infty$

$$\begin{aligned} \left| \int_{-\infty}^{+\infty} g(x) \sin(Nx) dx \right| &\leq \left| \int_{-\infty}^{+\infty} g(x) \sin(Nx) dx \right| + \int_b^{+\infty} |g(x)| dx \\ &\quad + \int_{-\infty}^a |g(x)| dx. \end{aligned}$$

Using here the previous results we obtain that for any $-\infty < a < b < +\infty$

$$\limsup_{N \rightarrow \infty} \left| \int_{-\infty}^{+\infty} g(x) \sin(Nx) dx \right| \leq \int_b^{+\infty} |g(x)| dx + \int_{-\infty}^a |g(x)| dx.$$

Passing here to limit as $b \rightarrow +\infty$ and $a \rightarrow -\infty$ we obtain the first limit in (A.4). Similarly, we can obtain the second one. \square

Let us now study the inverse Fourier transform. To this end, we need the following local Dini condition.

D) Assume that, for some fixed $x \in \mathbb{R}$, there exist the finite limits

$$g(x-) = \lim_{z \rightarrow x-} g(z) \quad \text{and} \quad g(x+) = \lim_{z \rightarrow x+} g(z)$$

and there exists $\delta = \delta(x) > 0$ for which

$$\int_0^\delta \frac{|g(x+t) + g(x-t) - g(x+) - g(x-)|}{t} dt < \infty.$$

Proposition A.5. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a function from $\mathbf{L}_1(\mathbb{R})$. If, for some $x \in \mathbb{R}$, this function satisfies the condition **D**, then

$$g(x+) + g(x-) = \frac{1}{\pi} \int_{\mathbb{R}} e^{-i\theta x} \widehat{g}(\theta) d\theta, \quad (\text{A.5})$$

where

$$\widehat{g}(\theta) = \int_{\mathbb{R}} e^{i\theta t} g(t) dt.$$

Proof. First, for any fixed $N > 0$ we set

$$J_N(x) = \frac{1}{2\pi} \int_{-N}^N e^{-i\theta x} \widehat{g}(\theta) d\theta = \frac{1}{\pi} \int_{\mathbb{R}} g(z) \int_0^N \cos(\theta(z-x)) d\theta dz,$$

i.e.,

$$J_N(x) = \frac{1}{\pi} \int_{\mathbb{R}} g(z) \frac{\sin(N(z-x))}{z-x} dz = \frac{1}{\pi} \int_0^\infty (g(x+t) + g(x-t)) \frac{\sin(Nt)}{t} dt.$$

Taking into account that for any $N > 0$ the integral

$$\frac{2}{\pi} \int_0^\infty \frac{\sin(Nt)}{t} dt = 1 \quad (\text{A.6})$$

and denoting $B(x) = (g(x+) + g(x-))/2$, we obtain that

$$J_N(x) - B(x) = \frac{1}{\pi} \int_0^\infty \frac{\omega(x, t) \sin(Nt)}{t} dt \quad \text{and} \quad \omega(x, t) = g(x+t) + g(x-t) - 2B(x).$$

Now we represent the last integral as

$$\int_0^\infty \frac{\omega(x, t) \sin(Nt)}{t} dt = I_{1,N} + I_{2,N} - 2B(x)I_{3,N},$$

where

$$I_{1,N} = \int_0^\delta \frac{\omega(x, t)}{t} \sin(Nt) dt, \quad I_{2,N} = \int_\delta^\infty G(t) \sin(Nt) dt, \quad I_{3,N} = \int_\delta^\infty \frac{\sin(Nt)}{t} dt$$

and $G(t) = (g(x+t) + g(x-t))/t$. Condition **D** and Lemma A.4 imply directly the convergence $I_{1,N} \rightarrow 0$ as $N \rightarrow \infty$. Now note that, since $g \in \mathbf{L}_1(\mathbb{R})$, then the function G is absolutely integrated. Therefore, in view of Lemma A.4, $I_{2,N} \rightarrow 0$ as $N \rightarrow \infty$. As to the last integral we use the property (A.6), i.e., the changing of the variables gives

$$I_{3,N} = \int_{\delta N}^\infty \frac{\sin t}{t} dt \rightarrow 0 \quad \text{as} \quad N \rightarrow \infty.$$

Hence we have the desired result. \square

References

- [1] O. E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial mathematics. *J. Royal Stat. Soc.*, **B 63**, 167–241, 2001.
- [2] V. S. Barbu and N. Limnios. *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications - Their use in Reliability and DNA Analysis*. Lecture Notes in Statistics, **191**, Springer, New York, 2008.
- [3] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall, 2004.
- [4] K. Bichteler and J. Jacod. *Calcul de Malliavin pour les diffusions avec sauts: existence d’une densité dans le cas unidimensionnel*. Séminaire de probabilité, XVII, Lecture Notes in Math., **986**, Springer, Berlin, 132–157, 1983.
- [5] C. M. Goldie. Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability*, **1** (1), 126–166, 1991.
- [6] I. A. Ibragimov and R. Z. Khasminskii. *Statistical Estimation: Asymptotic Theory*. Springer, Berlin-New York, 1981.
- [7] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. 2nd edition, Springer, Berlin, 2002.
- [8] R. Höpfner and Yu. A. Kutoyants. On LAN for parametrized continuous periodic signals in a time inhomogeneous diffusion. *Statist. Decisions*, **27** (4), 309–326, 2009.
- [9] R. Höpfner and Yu. A. Kutoyants. Estimating discontinuous periodic signals in a time inhomogeneous diffusion. *Statistical Inference for Stochastic Processes*, **13** (3), 193–230, 2010.
- [10] V. V. Konev and S. M. Pergamenshchikov. Sequential estimation of the parameters in a trigonometric regression model with the gaussian coloured noise. *Statistical Inference for Stochastic Processes*, **6**, 215–235, 2003.
- [11] V. V. Konev and S. M. Pergamenshchikov. Nonparametric estimation in a semimartingale regression model. Part 1. Oracle Inequalities. *Vestnik Tomskogo Universiteta, Mathematics and Mechanics*, **3**, 23–41, 2009.
- [12] V. V. Konev and S. M. Pergamenshchikov. Nonparametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency. *Vestnik Tomskogo Universiteta, Mathematics and Mechanics*, **4**, 31–45, 2009.

- [13] V. V. Konev and S. M. Pergamenshchikov. General model selection estimation of a periodic regression with a Gaussian noise. *Annals of the Institute of Statistical Mathematics*, **62**, 1083–1111, 2010.
- [14] V. V. Konev and S. M. Pergamenshchikov. Efficient robust nonparametric estimation in a semimartingale regression model. *Ann. Inst. Henri Poincaré Probab. Stat.*, **48** (4), 1217–1244, 2012.
- [15] V. V. Konev and S. M. Pergamenshchikov. Robust model selection for a semimartingale continuous time regression from discrete data. *Stochastic processes and their applications*, **125**, 294–326, 2015.
- [16] D. Lamberton and B. Lapeyre. *Introduction to Stochastic Calculus Applied to Finance*. Chapman & Hall, London, 1996.
- [17] N. Limnios and G. Oprisan. *Semi-Markov Processes and Reliability*. Birkhäuser, Boston, 2001.
- [18] R. Sh. Liptser and A. N. Shiryaev. *Theory of Martingales*. Springer, 1986.
- [19] T. Mikosch. *Non-Life Insurance Mathematics. An Introduction with Stochastic Processes*. Springer, 2004.
- [20] C. Mallows. Some comments on C_p . *Technometrics*, **15**, 661–675, 1973.
- [21] C. Marinelli and M. Röckner. On maximal inequalities for purely discontinuous martingales in infinite dimensions. *Séminaire de Probabilités, Lect. Notes Math.*, **XLVI**, 293–315, 2014.
- [22] M. Nussbaum. Spline smoothing in regression models and asymptotic efficiency in L_2 . *Ann. Statist.*, **13**, 984–997, 1985.
- [23] A. A. Novikov. On discontinuous martingales. *Theory Probab. Appl.*, **20** (1), 11–26, 1975.
- [24] M. S. Pinsker. Optimal filtration of square integrable signals in gaussian white noise. *Problems of transmission information*, **17**, 120–133, 1981.